C
10

ISSN 0037-8437

ANALES

DE LA

SOCIEDAD CIENTIFICA ARGENTINA

AÑO 2012 - VOLUMEN 246 - N° 1

MCZ
LIBRARY
JUL 07 2016
HARVARD
UNIVERSITY

SUMARIO

SOCIEDAD CIENTIFICA ARGENTINA

JUNTA DIRECTIVA 2011 - 2013

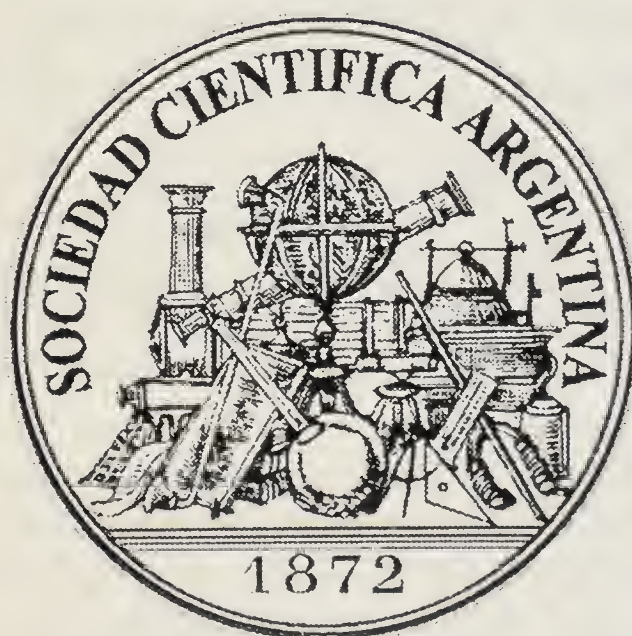
Presidente	Dr. Angel Alonso
Vicepresidente 1°	Dr. Eduardo Castro
Vicepresidente 2°	Ing. Juan José Sallaber
Secretario	Dr. Ernesto O. Celman
Prosecretario	Ing. Juan María Cardoni
Tesorero	Dr. Raúl E. Vaccaro
Bibliotecario	Prof. Lic. Norma I. Sanchez
Vocales Titulares	Dr. José L. Speroni
	Dr. Norberto C. Sarubinsky Graffin
	Dr. Horacio H. Camacho
	Dr. Pablo M. Jacovkis
	Lic. Mario Eduardo Laplagne
	Dr. Eduardo A. Pigretti
	Prof. Carlos Alberto Ríos
	Dr. Jorge R. Vanossi
	Dr. Pedro R. Yáñez
Vocales Suplentes	Dr. Carlos Azize
	Dr. Alejandro De Nicola
	Dr. Carlos de Jorge
	Dr. Arturo L. Otaño Sahores
	Dra. Georgina R. de Lores Arnaiz
	Ing. Enrique Draier
Revisores de Cuentas	Ing. Gerardo H. Foege
	Dr. Rodolfo P. Rothlin
Consejo de Honor	Dr. Augusto Belluscio
	Dr. Nicolás O. Breglia
	Dr. Alberto R. Dalla Vía
	Dr. Alberto Boveris
	Dr. Horacio J. Sanguinetti

ANALES

DE LA

SOCIEDAD CIENTIFICA ARGENTINA

AÑO 2012 - VOLUMEN 246 - Nº 1



MCZ
LIBRARY
JUL 0 / 2016
HARVARD
UNIVERSITY

Avda. SANTA FE 1145
C1059ABF BUENOS AIRES - ARGENTINA
Correo Electrónico: sociedad@cientifica.org.ar
www.cientifica.org.ar

EX PRESIDENTES DE LA SOCIEDAD CIENTIFICA ARGENTINA

1872-1874 Ing. Luis A. Huergo
1874-1875 Dr. Juan J. J. Kyle
1875-1877 Ing. Pedro Pico
1877-1878 Ing. Guillermo White
1878-1879 Ing. Luis A. Huergo
1879-1880 Dr. Valentín Balbín
1880-1881 Dr. Carlos Berg
1881-1882 Ing. Luis A. Huergo
1882-1883 Dr. Carlos Berg
1883-1885 Ing. Guillermo White
1885-1886 Ing. Luis A. Viglione
1886-1887 Dr. Estanislao S. Zeballos
1887-1889 Dr. Valentín Balbín
1889-1891 Dr. Carlos Maria Morales
1891-1892 Ing. Eduardo Aguirre
1892-1893 Dr. Juan J. J. Kyle
1893-1894 Ing. Carlos Bunge
1894-1895 Ing. Miguel Iturbe
1895-1896 Dr. Carlos Maria Morales
1896-1897 Dr. Angel Gallardo
1897-1898 Ing. Domingo Nocetti
1898-1900 Ing. Dr. Marcial R. Candiotti
1900-1901 Dr. Manuel B. Bahía
1901-1902 Dr. Carlos Maria Morales
1902-1903 Ing. Carlos Echagüe
1903-1904 Ing. Emilio Palacio
1904-1906 Dr. Carlos Maria Morales
1906-1908 Ing. Gral. Arturo M. Lugones
1908-1909 Ing. Otto Krause
1909-1910 Ing. Vicente Castro
1910-1911 Dr. Francisco P. Moreno
1911-1912 Ing. Vicente Castro
1912-1913 Gral. Dr. Agustín Alvarez
1913-1914 Ing. Santiago E. Barabino
1914-1915 Dr. Francisco P. Lavalle
1915-1917 Ing. Nicolás Besio Moreno
1917-1919 Dr. Carlos Maria Morales
1919-1923 Ing. Santiago E. Barabino
1923-1927 Ing. Eduardo Huergo
1927-1929 Ing. Nicolás Besio Moreno
1929-1933 Dr. Nicolás Lozano
1933-1937 Ing. Nicolás Besio Moreno
1937-1943 Ing. Jorge W. Dobranich
1943-1946 Dr. Gonzalo Bosch
1946-1949 Ing. José M. Páez
1949-1951 Ing. Dr. Eduardo María Huergo
1951-1953 Dr. Abel Sánchez Díaz
1953-1955 CERRADA
1955-1956 Dr. Abel Sánchez Díaz
1956-1959 Dr. Eduardo Braun Menéndez
1959-1962 Ing. Pedro Longhini
1962-1964 Dr. Pablo Negroni
1964-1970 Ing. José S. Gandolfo
1970-1976 C. de Nav. Emilio L. Díaz
1976-1988 Ing. Agr. Eduardo Pous Peña
1988-1989 Ing. Augusto L. Bacqué
1989-1992 Ing. Lucio R. Ballester
1993-1999 Dr. Arturo Otaño Sahores
1999-2001 Dr. Andrés O. M. Stoppani
2001-2005 Dr. Alfredo Kohn Loncarica
2005-2009 Dr. Jorge R. A. Vanossi

FIEBRE HEMORRAGICA ARGENTINA. OTRA MIRADA HISTÓRICA
(Ciencia y tecnología)

Norma Isabel Sánchez [1]
Sandra Janete Inwentarz [2]

[1] Profesora adjunta de Historia de la Medicina y directora el Observatorio de Historia de la Medicina, del Instituto y Cátedra de Historia de la Medicina. Departamento de Humanidades Médicas (FM/UBA). Ce: akohnlon@fmed.uba.ar

[2] Del Departamento de Neumonología. Instituto Profesor Doctor Raúl Vaccarezza. Cátedra de Medicina Respiratoria. Ce: sandrainwentarz@yahoo.com.ar
La co-autora expresa un sincero agradecimiento a la familia Barrera Oro por los datos suministrados. A Leónidas Asdrúbal Barrera Oro por las horas dedicadas a relatar lo sucedido en aquellos años difíciles, a mostrar algunos de los manuscritos de su hermano y anotaciones personales que atesora. A Gabriel F Barrera Oro por el tiempo que ocupó buscando datos, escritos y recortes de periódicos.

RESUMEN
La Fiebre Hemorrágica Argentina (FHA), es producida por un arenavirus conocido como virus Junín, que tiene la particularidad de establecer infecciones crónicas en un roedor, el Calomys musculinus, que transmite la enfermedad a través del contacto del hombre con su saliva, orina, sangre o sus partículas aerosolizadas. La zona endémica cubre aproximadamente 150.000 km^2, comprometiendo las provincias de Buenos Aires, Córdoba, Santa Fe, Entre Ríos y de La Pampa, con una población estimada en riesgo de 5 millones. Los primeros afectados se registran a partir del tramo final de la década de 1940; pero, sólo alrededor de 1955 toma mayor notoriedad, cuando el número de casos aumenta. Es entonces cuando un grupo de investigadores se dirigen hacia allí, tratando de paliar la situación. Sobresale entre ellos Julio Guido Barrera Oro, virólogo del Instituto Malbrán, quien ante la incertidumbre acerca del origen de esta enfermedad decide autoinocularse tratando de encontrar una respuesta. Después de variadas vicisitudes, todo culminó con el descubrimiento de una vacuna para proteger a la población zonal.

Palabras claves: Fiebre Hemorrágica Argentina; mal de los rastrojos; mal de O'Higgins; virus Junín; Candid 1; Julio Guido Barrera Oro.

BREVE DETALLE DE LA ENFERMEDAD

En estas páginas reseñaremos, de manera resumida, las características de la dolencia, de qué manera fue hallado el tratamiento (con plasma del enfermo convaleciente) y cómo se llegó a la fabricación de la vacuna, a partir de los relatos de los familiares de Julio Guido Barrera Oro, uno de los protagonistas de esta

historia.

La Fiebre Hemorrágica Argentina (FHA), localmente denominada "mal de los rastrojos" o "mal de O'Higgins", es una fiebre hemorrágica viral, zoonótica, causada por el virus Junín, un arenavirus, muy relacionado con el virus Machupo (agente causante de la fiebre hemorrágica boliviana). Su vector es una especie de roedor, la laucha del maíz o ratón maicero (Calomys musculinus), que lo desparrama a través de su saliva, orina o sangre y entra en contacto con el humano por la piel (a través de escoriaciones o heridas); las mucosas (conjuntival, oral: el clásico morder hebras de hierba o palitos); la inhalación de partículas aerosolizadas portadoras del virus.

La zona endémica cubre cerca de 150.000 km² comprometiendo las provincias de Buenos Aires, Córdoba, Santa Fe, Entre Ríos y de La Pampa, con una población estimada en riesgo de 5 millones. Se produce principalmente entre personas que residen, visitan o trabajan en el medio rural (más frecuente en este último caso); el 80 % de los infectados son hombres entre 15 y 60 años. De febrero a mayo de cada año, cuando el laboreo agrícola es intenso, pueden ocurrir miles de casos, con una tasa de mortalidad global de 10 a 25% [3].

Enfermedad aguda que comienza como un simple cuadro gripal, con un período de incubación de cerca de 10 días y los primeros síntomas que se presentan son: fiebre alta, malestar general, astenia, fotofobia, cefaleas, dolor retroocular, anorexia, náuseas, mareos. En el examen físico se observa un exantema eritematoso localizado en la cara, el cuello y el tronco, petequias en la piel, congestión generalizada de la mucosa oral y adenopatías[4]. Predominan los cuadros hemorrágicos con hematemesis, melena, hemoptisis y/o hematuria. En casos severos, aparecen síntomas neurológicos entre el 6° y 8° día, con ataxia, hipersomnia o insomnio, temblores especialmente de la lengua y convulsiones tónico-clónicas. Los fenómenos hemorrágicos tanto macro como microscópicos son constantes y se pueden ver en casi todos los órganos (bazo, pulmón, corazón, ganglios linfáticos, intestino, etc); la médula ósea muestra depleción de las 3 series, especialmente la eritroide. Todos estos fenómenos se producen tanto por agresión viral directa como por coagulación intravascular diseminada, manifestada por la presencia de trombos fibrinosos intravasculares. También por afección viral directa se origina una miocarditis intersticial con un infiltrado linfocitario entre las fibras miocárdicas, membranas hialinas pulmonares, necrosis de las papilas renales, cuerpos acidófilos hepáticos con necrosis hepática focal. Todo esto lleva al paciente a padecer procesos infecciosos secundarios, al shock irreversible, coma y muerte [5, 6].

La enfermedad también puede presentarse en forma más leve y, en ocasiones, el enfermo ignora que la ha padecido, pues tiene muchas similitudes con el síndrome gripal, con mandato de los síntomas gástricos y se podría hablar de una forma tífica, que, en caso de agravarse y predominar las formas vasculares, podría denominarse enfermedad tífico-vascular. Nunca la enfermedad es pura; en caso que domine la afectación del sistema nervioso, forma tífica-nerviosa. Es decir, se expresa bajo 3 formas: leve, moderadamente grave y grave, todas con manifestaciones bucales: hay enrojecimiento de las fauces, a veces un ribete eritematoso contornea los dientes; en la cara posterior de los labios y en la mucosa que tapiza la cara interna de las mejillas existe, con frecuencia, una red capilar rojo vinosa; se acompaña de piqueteado equimótico, más atenuado en la parte posterior de la boca, pilares anteriores, base de la úvula y paladar blando (en este aparecen 20 a 30 microvesículas de contenido claro). Este enantema bucal, signo importantísimo para el diagnostico diferencial, no está acompañado por ganglios de la región subángulo maxilar ni tampoco por odinofagia. Los pacientes tragan sin dificultad. La lengua que al principio es saburral, con la evolución de la enfermedad, se torna seca, aporcelanada ("lengua de loro" o "lengua tífica") y, al ser proyectada fuera de la boca, tiene un "temblor"[7], tanto más grosero cuando más grave es la evolución y está en relación con los síntomas encefaliticos que acompañan. La halitosis es muy acentuada y la gingivorragia está en todos los pacientes.

Las formas graves evolucionan hacia un importante porcentaje de desenlace fatal. Los enfermos presentan acentuado delirio, tics, convulsiones, gritos incoordinados, síntomas psicóticos y coma, seguido de muerte.

UN POCO DE HISTORIA

1943

Es posible que los primeros casos, de que se tenga noticias, sean de este año cuando un número más o menos significativo de pobladores de la zona noroeste de la provincia de Buenos Aires (partidos de 9 de Julio, Bragado, Junín, General Viamonte, Chacabuco, Rojas), padecieron "gripe" con fiebre muy alta y un 60% de mortalidad, llamada -por los lugareños- "la fiebre"[8].

1955

Momento inicial de la comunicación oficial, realizada por Rodolfo Arribálzaga[9], jefe de la Sala de Infecciosas del Hospital de Bragado. En su trabajo publicado por El Día Médico, del 16 de junio de 1955, describió una serie de casos que cursaban como un tipo de "gripe", con hipertermia, exantema, enantema, nefrotoxicidad y leucopenia y solicitó, por lo tanto, la presencia de especialistas de Buenos Aires para su resolución. Se supo posteriormente que sus reclamos no fueron verificados; en realidad era una pésima oportunidad para prestarle atención pues el país vivía momentos de incertidumbre con un nuevo intento de levantamiento militar que buscaba derrocar al presidente Perón que, meses más tarde, en efecto, se concretaría (setiembre de 1955), para dar paso a la denominada Revolución Libertadora.

1956-1957

Durante este período el país vive momentos de alta intolerancia política; los tres poderes del Estado están intervenidos y, en medio de situaciones de inestabilidad, poca trascendencia se brindó a ciertos temas, como el que nos ocupa. Sin embargo (como explicamos, líneas abajo, en otro apartado) se le dio forma al Instituto Nacional de Microbiología (INM), con tempranos laboratorios de biología molecular e Ignacio Pirosky va a cumplir una labor destacada, mientras incorpora investigadores de la talla de Julio Barrera Oro, Julio I Maiztegui, César y Celia Milstein, Armando S Parodi, Eugenia Sacerdote de Lustig, Juan A Zuccarini y muchos otros.

1958

El 1° de mayo Arturo Frondizi toma la conducción del país y hay un intento de mejorar o redireccionar situaciones anómalas. En su gabinete, Héctor V. Noblía es el responsable del área de salud.

Por entonces, muchos médicos zonales se enterarían de esta enfermedad, a través de los artículos publicados por el diario La Razón[10], [11], [12] y, ante la reiteración de casos, la zona afectada fue visitada por más de una comisión específica: una Médica Militar, otras de la UBA, del Instituto Nacional de Microbiología Carlos Malbrán y del instituto Biológico de la provincia de Buenos Aires. Un testigo de la época nos relató:

"Se tomó un nido de ratones maiceros y se lo trasladó a Buenos Aires. Se colocó en los mismos, lauchas lactantes que al ser picadas por los ácaros, enfermaron con características similares a los seres humanos enfermos. Además se inocularon animales de laboratorio con solución fisiológica que contenía ácaros machacados, enfermando también estos animales…

Sin embargo el ciclo roedor-ácaro-hombre está en pleno estudio …"[13].

Va a resultar un año muy significativo y entra en escena nuestro protagonista principal, Julio Barrera Oro, quien expresó oportunamente:

"El 21 de junio … fui a Junín, como integrante de la Comisión Nacional "ad hoc" nombrada por el Señor Ministro de Asistencia Social y Salud Pública, Dr Noblía, y formada por el jefe de ella, Dr Pirosky, y por los doctores Molinelli, Zuccarini y Di Pietro, que también tenían a su cargo tareas directrices; el Dr De León, que estudiaba la parte epidemiológica; el Dr Bernabé Ferreira, especialista en Virología; y el señor Teodomiro Vázquez, Entomólogo, jefe del equipo de campo que comandaba el grupo de desratizadores,

cuyo capataz, señor Díaz, tuvo la desgracia de enfermar allí del "Mal de los Rastrojos".
El Dr Pedro Martini, el estudiante de medicina F Gutman Frugoni y yo trabajábamos en el laboratorio, en todo lo concerniente a exámenes de sangre y otras muestras extraídas de los enfermos, buscando al agente patógeno de la enfermedad en cuestión ...
Quiero dejar constancia de mi agradecimiento al señor Francisco Gianelli, quien nos preparaba los materiales de trabajo, pues sin sus servicios, nos hubiera resultado dificultosa nuestra tarea ..."[14].

Líneas más abajo vamos a completar esta información. Eran meses en que se trabajaba en las zonas afectadas recopilando datos y todo material que resultase provechoso con el propósito de "entender" la enfermedad, aislar al agente etiológico y encontrar así un tratamiento efectivo.
Continúa diciendo:

"Ya en Buenos Aires, por razones circunstanciales, recayó en mi el honor de recibir en el Instituto Nacional de Microbiología al doctor Earl Chamberlain, consultor de la Oficina Sanitaria Panamericana ...
Al observar los resultados logrados hasta entonces por nuestro equipo, resumió su opinión en un ´Too Good´, que yo no supe bien si era admirativo o escéptico. Entonces repliqué al ilustre visitante que nuestras investigaciones no serían publicadas mientras no pudiera producirse la enfermedad en el hombre, experimentalmente ...
A comienzos de diciembre, las pruebas efectuadas por nuestro equipo habían permitido determinar ya, de un modo irrefutable, mediante el proceso de ultrafiltración, que habíamos aislado un virus, y los ensayos serológicos hacían suponer, con amplio márgen de seguridad, que ese virus era el agente causal del ´mal de los rastrojos´...
Faltaba la prueba más concluyente: provocar la enfermedad en el hombre, experimentalmente ...
Comencé a acariciar la idea de la autoinoculación ...Tenía muchas dudas al respecto ... las consecuencias sobre mi familia ... razones de disciplina ...
Comencé a hablar de ello con algunos compañeros y todos trataban de disuadirme. Solamente logré convencer de la necesidad de esta prueba a una camarada del equipo: la doctora Lidia Martos y ella prometió ayudarme en los primeros pasos: pero me traicionó, informó de esto al señor Zuccarini y yo me enteré. Todo esto ocurría el 19 de diciembre ...
Hice que me extrajeran sangre de la vena para saber en qué condiciones se encontraba mi organismo, si tenía defensas orgánicas, etc[15].

Dejemos por ahora en suspenso este relato y prosigamos con la narrativa general.
El 19 de diciembre, en el aula magna de la Facultad de Medicina (UBA), se realiza un encuentro científico, oportunidad en la que se presenta todo lo que se imputa como interesante. Se visualizan dos posiciones: las comisiones de la Facultad de Medicina y del Instituto Nacional de Microbiología, sustentaban con firmeza la teoría de la naturaleza viral de la enfermedad; en tanto, otros investigadores persistían en la teoría de la Leptospira, una bacteria, como agente etiológico. No faltaron aquellos que sostenían estaba relacionada a procesos tóxicos determinados por las fumigaciones en los campos con (el insecticida) dieldrin para combatir las tucuras[16].
Ante tal decepcionante situación, Barrera Oro toma la decisión de auto-inocularse y redacta un informe para la lectura de sus potenciales continuadores[17], pues no tiene la certeza total de los resultados.
Nos detenemos, una vez más, para hacernos dos preguntas: por un lado una de carácter metodológico, ¿era el momento oportuno para llevar adelante esta "inoculación"?
La otra desde una perspectiva ética, ¿era correcta o por lo menos aceptable tan personal y, posiblemente, precipitada decisión? Un entendido, tal vez, nos diría que lo guió el principio de benevolencia: se dispuso a actuar en beneficio de otros. Y, a la vez, el ideal de beneficencia, que implica actos supererogatorios, que este caso particular fue asumir una actitud altruista.
El acto de inoculación que se auto-indicó Barrera Oro merece algunas reflexiones. Antes, es preciso comentar que tales actitudes frente a un problema científico son poco frecuentes y forman parte de un tiempo

pasado, incluso anterior a la época en que actuó este investigador. Ejemplos, con algún grado de similitud, aparecen en Edward Jenner (al inocular a su hijo con lo que luego sería la vacuna antivariólica) y en Daniel Alcidez Carrión García (quien lo hizo con un extracto de verruga peruana con consecuencias nefastas: posteriormente, se dispuso que la enfermedad llevase su nombre).
Tales conductas, por un lado, muestran un grado de decisión mezclado con el deseo de llegar a la verdad y, por otro, hay una cuestión relacionada con los tiempos de espera. En efecto, al presente, los modelos animales proveen de técnicas con las que es posible identificar si un microbio es el agente etiológico (virus por ejemplo) de una enfermedad o no. Asimismo ocurre con drogas, vacunas y demás elementos que luego van a ser usados en el ser humano. La generación de modelos animales patológicos y su validación insume tiempo que se puede medir en meses o años.
En el caso de Barrera Oro, es importante dejar en claro que aún en el caso que al inocularse hubiera dudas sobre el rol del virus involucrado, la investigación no se iba a detener allí. Por el contrario, la generación de terapéuticas, tales como una vacuna, siempre iban a precisar de modelos animales patológicos.
Volvamos al relato del protagonista, en primera persona:

"De la Facultad salí totalmente decidido a inocularme y me vine al Instituto. Lo primero que hice fue sacrificar tres ratones enfermos, les saqué el cerebro, preparé una suspensión de Hanks con antibiótico e hice los controles de esterilidad de la suspensión, sembrándolos en medio de cultivo, y comencé el primer paso de la prueba de titulación que en el futuro, determinaría la cantidad de virus que me había inoculado ...
Con vergüenza declaro que no me animaba a darme yo mismo el pinchazo ... llamé al sereno del Instituto Héctor Omán, que era estudiante de medicina y con el engaño que necesitaba darme una vacuna, le rogué que me inyectara en el hombro derecho, en la región deltoidea ...
Redacté luego las instrucciones para mis compañeros de equipo:
El objeto fundamental de este trabajo, es tratar de demostrar, definitivamente, que el agente patógeno aislado de los enfermos de Junín reproduce la enfermedad en el hombre, después de haber sido mantenido ese agente por pasajes en lauchas lactantes. Por lo tanto, se deberá prestar especial atención a la ejecución de las pruebas tendientes a demostrar que la enfermedad provocada experimentalmente, es la misma que afectó a los pacientes de Junín. Con este fin se dará prioridad a lo siguiente:
1°.- Inoculación de sangre exclusivamente, a lauchas lactantes (IC-SC-IN-IP), a cobayos (IP), a embrión.
2°.- Cultivos en Eugon broth, en medio con tioglicato y en medio de Vervoort (de sangre exclusivamente).
3°.- Recuento de glóbulos blancos.
4°.- Recuento de plaquetas.
5°.- Recolección de sangre, sistemáticamente todos los días después de aparecidos los síntomas de la enfermedad experimental, para pruebas serológicas.

Si se logra aislar supuestos agentes patógenos, se hará:
1°, antígenos para fijación de complemento que se ensayará con el suero 200 (está en la congeladora de la heladera de la sección Peste);
2°, pruebas de neutralización con los sueros 167 y 200.

Si el paciente no fallece, con sus sueros del periodo de convalecencia se hará:
1°, prueba de fijación de complemento frente a antígeno preparado con cepa Davio;
2°, prueba de neutralización con cepa Davio.
SE ADVIERTE SEVERAMENTE que no debe cometerse el error de tratar de estudiar todos los aspectos de la enfermedad experimental.
Acordarse siempre del objeto primordial del trabajo.
PARA CUMPLIR EL OBJETO DEL TRABAJO SE DEBERAN OBSERVAR ESTRICTAMENTE LAS SIGUIENTES INSTRUCCIONES:

1°) Todos los días, a partir del siguiente a la inoculación, se extraerá estérilmente 10 ml de sangre venosa heparinizada (aguja de platino); de esos diez centímetros cúbicos, uno estará destinado para lauchas (inocular una camada por día); y cuatro para cultivos (un matraz de Eugon broth, cinco tubos de tioglicolato, y cinco tubo de Vervoort por día). La extracción de sangre heparinizada se suspenderá en la mitad de la convalecencia.
2°) Todos los días, a partir del comienzo de los síntomas, se extraerán 10 ml de sangre y se conservará el suero a -20°c para pruebas serológicas.
3°) Las pruebas generales de laboratorio sólo se harán en 5 días de la totalidad del experimento: en un día del período de comienzo, en un día del período de estado, en un día sea del premortem, sea del comienzo de la convalecencia, en un día de la convalecencia.
RECORDAR SIEMPRE EL OBJETO PRIMORDIAL DEL TRABAJO.
EVITAR PERDERSE EN LOS LABERINTOS DEL LABORATORIO CLINICO.
4°) Si se declara la enfermedad, SE ORDENA DEJAR QUE SIGA SU EVOLUCION NATURAL. SE PROHIBE USAR SUERO DE CONVALESCIENTES. SE PROHIBE USAR ANTIBIOTICOS, (excepto en el caso de que aparezcan complicaciones bacterianas, diagnosticadas por cultivo). SE PERMITE USAR CUALQUIER RECURSO INESPECIFICO, QUE EL MEDICO CONSIDERE NECESARIO.
(Ruego a los que intervengan en este trabajo, que no se dejen influir por consideraciones sentimentales y cumplan fielmente el punto 4°)"[18].

Para completar tan interesante referencia, agregamos la explicación que, oportunamente, ofreciera el odontólogo Leónidas Barrera Oro, quien comentó que, en un primer momento, no tomó en serio las palabras de su hermano Julio Guido. Pero, el 21 de diciembre, a 48 horas de aquel episodio, le observa en la región deltoidea del brazo derecho una reacción intensa en la zona de inoculación y, a partir de allí, se concentra en la situación y así podrá relatar:

"Efectúa un preparado con 800 mg de células de laucha lactantes, inoculadas con la cepa "Davio" (en el pasaje13°), con 3.5 cc de liquido de Hanks más 0.5 de solución acuosa conteniendo 1000 U de Penicilina g Sódica y 500 mg de Sulfato de Estreptomicina.
El período de incubación de la enfermedad experimental fue breve, de 39 horas aproximadamente. El 23 de diciembre comienza con fiebre por la tarde; el 24, presenta cara "vultuosa y enrojecida": facies de ebrio, piel de la cara como expuesta a quemaduras de sol, la marcha era oscilante como de pato, la conjuntiva inyectada en sangre, fiebre no muy alta, de 38° aproximadamente, la mucosa de los labios presentaba un color gris que contrastaba con el gran enrojecimiento de las fauces, las encías presentaban un color rojo violáceo ...
El día 25 es internado en el Hospital Británico. Presentaba gran agitación, era difícil tenerlo en la cama, dolores en las pantorrillas y en la región lumbar, hiperestesias cutáneas, casi no soportaba las sábanas, temblor de lengua y epigastralgia que lamentablemente se fue agravando por la administración de "Corifedrine" que le administraban para aliviar los otros síntomas ...
Se lo sometió a un baño de inmersión caliente, que pareció calmarlo. Lo obligaron a ingerir líquidos, ya que rechazaba todo.
El día 26, comienza a establecerse la oliguria; el 28 aparecen las primeras epistaxis, después la hemoptisis. A esta altura tenía delirios y convulsiones.
La mejoría comienza a fines de diciembre. No puedo precisar bien la fecha, para el 30 ó 31. Como en muchos pacientes comienza con una gran poliuria, cesan al parecer los trastornos gástricos; comienza a ingerir alimentos, pero esto es seguido de vómitos, cesa la fiebre; comienza a recuperar la tensión arterial ... El 2 de enero regresa a su domicilio. La mirada resulta extraña, hasta mucho tiempo después pierde el cabello. Permanece en reposo varios días más" [19].

Esta descripción, realizada por el expreso pedido del paciente en cuestión, coincide en rasgos generales con

la forma de presentación de la enfermedad en las personas clasificadas como afectadas por FHA moderadamente grave.
Las pruebas de laboratorio, efectuadas por sus compañeros del equipo del INM permitieron comprobar la existencia en sangre del virus en el 3°, 6°, 7°, 9, 10° y 12° día desde la inoculación experimental, leucopenia con aneosinofilia y plaquetopenia, 130.000 plaquetas el 9° día, confirmando todo esto que el germen aislado era el causante de la enfermedad y que por sus características de comportamiento, era viral (desechando la teoría de la Leptospira y de los agentes tóxicos).

"A la semana de la autoinoculación, aproximadamente -continuamos leyendo-, un joven colega y discípulo de Julio Guido, el Dr Martini, no recuerdo el nombre[20], realizó una autoinoculación similar, quería seguir junto a su colega pero fue terrible ...
Comenzó rápidamente con una enfermedad extremadamente grave que lo llevó a una coagulación intravascular diseminada y falleció a las 48/72 horas ...
Nunca supimos bien si el inoculo fue superior al de Julio Guido o la respuesta del organismo fue diferente, fue terrible, se perdió un gran investigador ..."[21].

Después de 1959

Barrera Oro hizo su recuperación; mientras Pirosky, en los laboratorios del INM, procedía a aislar el virus. A la par, en la cátedra de Microbiología, de la FM/UBA, el investigador Armando S Parodi [22] y su equipo hacían lo propio y preparaba una vacuna que se empleó durante un tiempo. No fueron los únicos ya que unos cuantos más también estaban concentrados en la dolencia. Las motivaciones eran múltiples, donde no faltaban las económicas y políticas, ya sea porque afectaba a una zona de alta producción cerealera y ganadera, con demanda de mano de obra estacional y permanente, con commodities de altos rindes; ya sea de razones políticas, pues el gobierno democrático de la época estaba muy amenazado por fuerzas dispuestas a su clausura y buscaba pretextos para lograr su intento, justificándolo en razones de inoperancia.
Nacía el Conicet y, algo más tarde, la carrera de investigador científico.

1962

Una nueva interrupción democrática y las Fuerzas Armadas ubican en el Poder Ejecutivo Nacional, según la ley de acefalía, al presidente provisional del Senado: José M Guido, quien tendrá como asesor en el Ministerio de Salud a Tiburcio Padilla. El Instituto que conducía Pirosky fue intervenido y comienzan años de incertidumbre, mientras decae el interés (científico/político) por la enfermedad, no por cuenta de los posibles afectados y la población lugareña en su conjunto. Parte de los pesquisadores allí reclutados abandona el país o la investigación.

1964

Para entonces ya se sabía que la dolencia tenía relaciones con afecciones similares aparecidas en la India, China, Bolivia, Rusia y se llevaron a delante prolijas compulsas, más serios cuidados asistenciales y preventivos. Mientras el Congreso de la Nación aprueba la creación de la Comisión Nacional Coordinadora de Estudio y Lucha contra la FHA[23].

1965

Se instala en Pergamino un nuevo grupo de estudio, que suma esfuerzos a los del INM; entre ellos Julio Maiztegui[24], quien está acompañado por miembros el Centro de Educación Médica e Investigaciones Clínicas (CEMIC), quien dispuso de admiradores y detractores (como suele ser común en estas situaciones), con opiniones de idóneos y legos, de interesados y desinteresados.

1966

Llega el golpe militar del 28 de junio, encabezado por Juan C. Onganía, quien desalojó al presidente Arturo Illia y abortó, entre muchos otros, los esfuerzos que hacía este grupo humano arriba mencionado.

Ante la grave situación, que sumó a cuestiones personales, Julio G Barrera Oro parte a EEUU, donde continúa con sus investigaciones.

1971-1975

Maiztegui comenzó a tratar a los infectados recientemente con sueros de pacientes recuperados y redujo, de manera notable, la mortalidad (se calcula que al 1%). Hubo ensayos, vacunaciones, críticas, auspicios, avances, retrocesos[25].

1976 (y siguientes)

En marzo la sexta interrupción jurídico-política del siglo XX que afectó al país. Ahora la conducción es responsabilidad de una Junta Militar que interviene todos los poderes y hace todo tipo de re-designaciones.
Llega 1978 y nace el Instituto Nacional de Estudios sobre Virosis Hemorrágicas, en Pergamino (hoy: Instituto Nacional de Enfermedades Virales Humanas 'Julio Maiztegui', INEVH) para continuar con los estudios.
Por un convenio entre el INEVH, la OPS y un programa de la ONU, Barrera Oro fue comisionado para buscar una vacuna a virus vivos atenuados y viajó, otra vez, a EEUU. En ese tiempo trabaja en Fort Detrick (Maryland), en una instalación que pertenece al Comando Médico del Ejército de los EEUU: el Instituto de Investigación Médica sobre Enfermedades Infecciosas (Usamriid).

1981-1989

Barrera Oro aísla la cepa Candid 1[26] y el resultado fue la vacuna, altamente inmunológica, que fue cedida al INEVH.
Los ensayos clínicos de fases I, II y III, realizados entre 1985 y 1989, demostraban que la vacuna es segura y su eficacia en hombres de 15 a 65 años es del 95,5%. Además, se ha constatado, mediante anticuerpos neutralizantes, que la respuesta inmunitaria específica se mantiene hasta 13 años después de la vacunación en más de 90% de los vacunados [27]. Sin embargo, no hay interés comercial en su producción debido a que es considerada una vacuna "huérfana", sólo sirve para aplicar a número reducido de personas de una región de la Argentina[28].
Los diarios mantenían informada a la población. En una nota titulada: Etapa final de la vacuna elaborada por un científico argentino, leemos, casi al término del relato:

"Barrera Oro adelantó que el año próximo en un congreso a celebrarse en Mar del Plata, con la presencia de virólogos de jerarquía internacional, se darán a conocer oficialmente los resultados obtenidos mediante la vacunación humana".

1990 y siguientes

Se estuvo en condiciones de decir que la Candid 1 era efectiva. Los investigadores y múltiples colaboradores habían actuado en momentos de dictadura y democracia: por suerte, ahora parecía que se entraba de manera definitiva por esta senda.
Una vez más, varias notas en los diarios. Veamos algunos titulares: Vacuna para el Mal de los Rastrojos. Lo anunció Julio Maiztegui. Comprueban la elevada eficacia de una vacuna contra la fiebre hemorrágica[30].
Otro: Pergamino neutralizó la fiebre hemorrágica; en este aparece el siguiente resumen:

"la licenciada María Rosa Feuillade, a cargo del Instituto de Virosis Hemorrágica, explicó importantes detalles del descubrimiento.
La investigación a cargo de los doctores Julio Maiztegui y Julio Barrera Oro, empezó a fines de 1979 (sic) ...".
La vacuna, que no tiene efectos colaterales y genera anticuerpos, tendría costos de producción muy altos, aún difíciles de estimar ...
... esta ciudad está construyendo su propio laboratorio de alta seguridad, pero para que sea terminado se

necesitan los fondos prometidos por el Ministerio de Salud y Acción Social"[31] .

El reporte (de divulgación, no de carácter científico) como se puede ver, tiene algunas licencias.
Muy valiosas resultan las palabras de Maiztegui, quien dijo, en 1993, al recibir un premio:

"Me veo obligado a compartir ... la historia del mal de rastrojos que comienza en la República Argentina allá por 1955, y no son otros que Armando Parodi de la cátedra de Epidemiología y Julio Barrera Oro, el creador de la vacuna Candid 1"[32].

Años más tarde, aparecía una nota inquietante: Queda poca vacuna par la fiebre hemorrágica, que continuaba así:

"Problema: se trata de un producto elaborado en los Estados Unidos por cuya continuidad se teme; se experimentó con éxito una fórmula local, pero faltan fondos para fabricarla.
... según dijo la directora, Delia Enría y la investigadora Ana Ambrosio, los habitantes de la zona endémica quedarían expuestos al contagio, hoy prevenido por 200 mil dosis producidas en los Estados Unidos y en vías de agotarse debido a la campaña de vacunación que se extiende ...
En tanto, y como parte de un convenio internacional, el virólogo argentino Julio Barrera Oro viajó a los Estados Unidos donde desarrolló una semilla de vacuna, al tiempo que en Pergamino se inició la construcción de un edificio que cumpliera con las normas de seguridad biológica ...
Como la enfermedad es exclusiva de nuestro país y en el extranjero no se prevé seguir produciendo el antídoto, cuando se agoten las dosis existentes habrá 3.500.000 personas expuestas a contraer el mal ..."[33].

Valoramos el esfuerzo del periodismo. De esa manera se tiene la atención del público y, fundamentalmente, de las autoridades que, ante la presión ciudadana, encuentran la manera de sortear los obstáculos.

2000-2010
La ANMAT (Administración Nacional de Medicamentos, Alimentos y Tecnología Médica, creada en 1992, como dependencia del Ministerio de Salud y Acción Social), habilitó la planta, en Pergamino, para la producción de la vacuna[34]. Leemos:
"Existe un plan de vacunación, con miras a una gran reducción de casos y la vacuna ha demostrado también ser efectiva contra el virus Machupo y, por ende, considerada como tratamiento alternativo para la fiebre hemorrágica boliviana[35].

El resto puede completarse con la lectura del Anexo.

EL INSTITUTO NACIONAL DE MICROBIOLOGÍA (y sus sucesivos cambios)
Tal vez resulte innecesario hacer mención a éste, pues se ha escrito mucho y bien. Aún así, para algún lector menos informado, incluiremos algunas referencias.
Existió, desde los finales de la década de 1880, un Instituto Bacteriológico Nacional (como dependencia del Departamento Nacional de Higiene -DNH-) que contó con el apoyo de Carlos Malbrán y, poco después, de José Penna, profesional que impulsó la venida al país, hacia 1916, del importante bacteriólogo Rudolf Kraus (su director entre 1913 y 1921). Con él colaboraron o se formaron algunos de los más sobresalientes investigadores argentinos, al modo de Alfredo Sordelli y Bernardo A Houssay o Salvador Mazza y Ángel Roffo y fue un punto de arranque de nuestra tradición biomédica.

Con la llegada del peronismo, ese DNH se transformó en un ministerio y hubo modificaciones en su organigrama; en el primer postperonismo, el propio instituto fue remozado y aparece el primer laboratorio de biología molecular, enmarcado en un proceso más amplio de "modernización" en ciencia y tecnología, con la dirección de Ignacio Pirosky. En efecto, en 1956, asumió como director (por decisión del ministro de

Asistencia Social y Salud Pública, Francisco Martínez, en época de la Revolución Libertadora). El viejo Instituto Bacteriológico se transformó en Instituto Nacional de Microbiología "Carlos Malbrán"[36]. Un tiempo pasó hasta que Héctor V Noblía (ministro de Arturo Frondizi) ante un proceso infeccioso que afectaba a una zona de Buenos Aires (que es el leit motiv de este trabajo), decide instalar en Junín (junio de 1958 a setiembre de 1959) una Comisión Nacional ad hoc (que integran especialistas como el citado Pirosky, Juan A Zuccarini, Ernesto Molinelli, Arturo Di Pietro, Pedro Martini, Julio Barrera Oro y la asistencia técnica de Luis F Frugoni Gutman, Manuel A De León y personal auxiliar, más Lidia Martos, Matiel de d´Ampaire y Alberto F Pfeifer).
A Barrera Oro, Martini y Frugoni Gutman (y el propio Pirosky), se les asignó:

"Todo lo relativo a experiencias en animales de laboratorio, en particular en ratones, previo análisis y formulación del protocolo experimental"[37].

Seguimos leyendo:
"Primera inoculación experimental en el hombre. Para cumplir también en esta infectopatía los clásicos postulados de la microbiología, uno de los autores (J G Barrera Oro), por propia decisión, resolvió practicar en si mismo la primera reproducción experimental de la enfermedad, inoculándose el agente patógeno aislado ...
Conclusiones. Sobre la base del estudio anatomoclínico, de los datos epidemiológicos y de las investigaciones etiológicas, puede afirmarse: ...
5) La inoculación experimental al hombre ha permitido reproducir típicamente los cuadros clínico, hematológico y urinario de la enfermedad natural habiéndose recuperado el agente etiológico específico de la sangre circulante del voluntario.
6) De ácaros hematófagos, del suborden de los Mesostigmata, capturados en las áreas endemoepidémicas, se ha recuperado el virus causal"[38].

Este período de Pirosky (abril de 1956 a abril de 1962), con la conformación de dos nuevas secciones dentro del Instituto, de genética bacteriana y de biología molecular, concluye cuando es intervenido por el Ministerio de Salud Pública[39], tras la destitución del titular del Poder Ejecutivo Nacional. Hubo un período perdido que llega hasta la década de 1970[40].

BREVE DATOS SOBRE BARRERA ORO

Julio G fue incorporado al Instituto Malbrán en los concursos de 1957, con 28 años, donde permaneció hasta su retiro a EEUU. En este país realizó estudios posdoctorales en la Escuela de Medicina de la Baylor University (Houston/Texas).
Durante su etapa de trabajo en Maryland, en el Instituto de Investigación Médica sobre Enfermedades Infecciosas (Usamriid) desarrolló la vacuna Candid I.
En este logro hay unas cuantas figuras cardinales: Barrera Oro, Pirosky, Parodi, Maiztegui y muchos silenciosos colaboradores, que deben ser más de un centenar[41].
Recomendamos leer los fundamentos del Proyecto de Resolución de la H. Cámara de Diputados de la Nación (incluido en el anexo) que resultan muy esclarecedores.

CONCLUSIONES

La FHA es una enfermedad viral, de características regionales, muy grave ya que puede llevar a la muerte.
Fue Julio Guido Barrera Oro una pieza clave en la búsqueda de la solución, con resultados benéficos para muchos pobladores de una determinada región de la Argentina.
Se entrecruzan acá (y, no es una novedad, pues por lo general se puede hacer extensible a una multiplicidad de investigaciones) cuestiones humanas, sociales, económicas, políticas, antropológicas, etc.
No olvidemos las problemáticas del "poder" y, en este sentido analícese que fue el Ejército de EEUU quien

"prestó" su concurso para la investigación.

ANEXO

H. Cámara de Diputados de la Nación

PROYECTO DE RESOLUCIÓN

Texto facilitado por los firmantes del proyecto. Debe tenerse en cuenta que solamente podrá ser tenido por auténtico el texto publicado en el respectivo Trámite Parlamentario, editado por la Imprenta del Congreso de la Nación.

Nº de Expediente	0105-D-2007
Trámite Parlamentario	002 (02/03/2007)
Sumario	EXPRESION DE BENEPLACITO POR LA PRODUCCION DE LA PRIMERA VACUNA ARGENTINA CONTRA LA FIEBRE HEMORRAGICA "CANDID 1"
Firmantes	INGRAM, RODDY ERNESTO - WEST, MARIANO FEDERICO.
Giro a Comisiones	ACCION SOCIAL Y SALUD PÚBLICA

La Cámara de Diputados de la Nación

RESUELVE

Expresar su especial beneplácito por la producción de la primera vacuna argentina contra la fiebre hemorrágica -una zoonosis que puede ser mortal si no es tratada a tiempo y cuya población en riesgo se estima en 5 millones de personas- que se conocerá como "Candid 1" y será desarrollada en el Instituto Nacional de Enfermedades Virales Humanas "Julio Maiztegui", de Pergamino.

FUNDAMENTOS

Señor presidente:

La vacuna contra la fiebre hemorrágica, una zoonosis que puede ser mortal si no es tratada a tiempo y cuya población en riesgo se estima en 5 millones de personas, se producirá en el país a partir de un convenio firmado en la Casa Rosada.

El presidente Néstor Kirchner y el ministro de Salud, Ginés González García, anunciaron en Casa de Gobierno la fabricación nacional de la primera vacuna contra la fiebre hemorrágica argentina (FHA). La producción local de la droga, que se conocerá como "Candid 1", se llevará a cabo en el Instituto Nacional de Enfermedades Virales Humanas 'Julio Maiztegui' (INEVH), dependiente del Instituto Malbrán.

El "Maiztegui" es el primer laboratorio en la Argentina habilitado para la producción de vacunas virales humanas con certificación del cumplimiento de las normas internacionales de "Buenas Prácticas de Manufactura".

La Fiebre Hemorrágica es una enfermedad viral aguda grave, caracterizada por un síndrome febril con alteraciones hematológicas, neurológicas, renales y cardiovasculares, que puede producir la muerte si no es atendida.

El área endémica de la fiebre cubre aproximadamente unos 150 mil kilómetros cuadrados en las provincias

de Buenos Aires, Córdoba, Santa Fe y La Pampa, con una población de riesgo estimada en 5 millones de habitantes.

Los primeros casos de Fiebre Hemorrágica en el país se conocieron a principios de la década de 1950 y para 1958, se identificó su agente etiológico al que se denominó "virus Junín". La enfermedad es una "zoonosis con reservorio animal": se incuba en el roedor "Calomys musculinus" y se transmite a los humanos por contacto.

Actualmente, los pacientes disponen de un tratamiento específico denominado "plasma de convalecientes de la FHA", cuya administración precoz es eficaz y disminuye la mortalidad al 1%.

Hace 16 años que se utiliza una vacuna contra la Fiebre Hemorrágica de fabricación extranjera, eficaz en un 95.5% de los casos; la producción nacional de la droga es un proyecto tecnológicamente único en su tipo en la Argentina.

La metodología conducente a la producción, control y aseguramiento de calidad de una vacuna viva atenuada para uso humano, según normas internacionales de BPM, no ha sido abordada hasta la fecha por ninguna otra institución pública o privada en la Argentina.

Por "las dificultades para la producción de ciertas vacunas en el ámbito internacional, con desaparición de laboratorios productores comerciales debido a cuestiones de costo-beneficio", el Ministerio de Salud "adoptó una política de costo-efectividad priorizando valor a la vida humana".

Así, el impacto del proyecto dentro del Programa Nacional de Lucha contra la Fiebre Hemorrágica Argentina es sustancial, ya que permitirá abordar definitivamente el control de esta endemia y abriría la posibilidad de que la capacidad generada para su producción se pueda utilizar para el desarrollo de otras vacunas similares, según las necesidades del país.

Adema, la vacuna producida en la Argentina posicionará a nuestro país como productor de vacunas virales creadas completamente en el país.

La vacuna fue producida por primera vez en Estados Unidos. Entre 1985 y 1988 se inocularon más de 300 voluntarios humanos entre los que no se observaron efectos secundarios severos atribuibles a su aplicación. Asimismo en más del 90% de los voluntarios se detectaron anticuerpos contra el virus Junín. Ahora es necesario demostrar que la vacuna de producción local tiene exactamente la misma cantidad de anticuerpos a la producida en EEUU.

De ser así, la Candid I, versión argentina, será utilizada oficialmente en las próximas campañas de vacunación masiva para prevenir la enfermedad. Así se estaría ampliando la aplicación de la misma, lo que revertiría la condición de "excluidos" en la que estaban gran cantidad de mujeres, niños y familiares de trabajadores agrícolas como así también gente que no está directamente vinculada a la actividad rural.

Se estima que el estudio comenzará a realizarse a comienzos de 2005, una vez que haya sido reclutada la cantidad necesaria de personas. De un total de 950 voluntarios, estos serán divididos en dos grupos, uno de ellos recibirá la vacuna fabricada en Estados Unidos y el otro grupo, la fabricada en el INEVH. Ni los voluntarios ni los investigadores conocerán lo que han recibido las diferentes personas hasta los seis meses posteriores a la inmunización. El estudio se realizará ante los veedores de la comisión de Bioética de la provincia de Córdoba y los voluntarios deberán firmar un acuerdo de consentimiento y concurrir al Instituto Maiztegui (Pergamino) donde se realizarán las pruebas y los estudios de rigor. Tras los seis meses, se procederá a develar las identidades y el tipo de inoculación recibida por cada uno y se estudiarán los resultados

con el objetivo de obtener el mayor grado de aproximación y exactitud en cuanto a los anticuerpos presentes en una y otra vacuna.

La enfermedad se da sólo en la Argentina y que afecta aproximadamente a 5 millones de personas en la zona endémica, es decir, aquella en la que se han registrado casos está compuesta por el sur de la provincia de Santa Fe, Córdoba, norte de Buenos Aires y La Pampa. Las investigaciones en las que participaron 6500 voluntarios en el sur de Santa Fe entre 1988 y 1990 permitieron comprobar que la Candid 1 protege al ser humano contra la enfermedad con una eficacia del 95.5%.

En base a estos estudios, entre 1991 y 2000 se procedió a vacunar a 227.994 pobladores de diferentes localidades de la provincia de Santa Fe, Córdoba y Buenos Aires. La efectividad de la vacuna bajo este protocolo es del 98%, sin embargo se ha registrado una extensión del área endémica, lo que aumenta la cantidad de población en riesgo a 5 millones de personas.

También se la conoce como el mal de los rastrojos, es causada por el virus Junín y no es erradicable ya que su reservorio son los ratones. La prevención puede hacerse mediante la administración de plasma inmune, es decir, plasma proveniente de personas que tuvieron Fiebre Hemorrágica Argentina y se recuperaron. Este plasma contiene sustancias llamadas anticuerpos, los cuales son capaces de actuar contra el virus Junín. Los más vulnerables son las personas que viven en el campo y en caso de no ser atendidas adecuadamente, la patología puede provocar un 30% de mortalidad.

En 1971 el Dr Julio Maiztegui investigó la Fiebre Hemorrágica y demostró que la mortalidad del 30% que presentaba la enfermedad se reducía al 3% si los pacientes eran tratados con plasma de personas ya enfermas antes del 8º día de haber contraído el mal.

Si bien la aplicación sistemática de plasma se inició en 1986, Maiztegui se desveló para conseguir una vacuna efectiva. Los esfuerzos continuos a pesar de la falta de recursos permitieron desarrollar la etapa experimental de la vacuna cuya viabilidad fue probada en voluntarios en la Ciudad de Santa Fe.
La vacuna Candid 1 está categorizada internacionalmente como "droga huérfana", es decir un producto imprescindible para una población restringida, cuyo nivel de demanda no la hace comercialmente interesante. Esto significa que a los laboratorios no les conviene producir una vacuna que solo será utilizada por un sector del país. Por eso las inmunizaciones serán financiadas por el Estado y serán gratuitas para el público.
Hoy es un logro para el INEVH haber transitado todas las etapas necesarias para la creación de la vacuna y que implica cumplir con los requisitos de bioseguridad.

Finalmente el sueño de producir en forma local la dosis contra la FHA del Dr Maiztegui, quien comenzó sus investigaciones allá por el año 1950 y demostró la viabilidad de la vacuna Candid 1 que luego se produjo en Estados Unidos, está más cerca de convertirse en una realidad.

Consideramos oportuno reconocer esta decisión de contribuir al mejoramiento de la salud humana, al avance científico y tecnológico y a la formación de recursos humanos por medio del desarrollo de medicamentos. La mejora en la salud de la población que trae aparejada la producción de una vacuna como la mencionada, junto a los estudios fármaco-epidemiológicos que a este respecto se realizan, obra en servicio de la toma de decisiones que planteen soluciones posibles a las enfermedades argentinas de difícil control.

Por los fundamentos expuestos solicito a mis pares la aprobación del presente proyecto de resolución.
Instituto Superior de Enfermedades Virales "Dr Julio I Maiztegui"
Monteagudo 2510-2700
Pergamino, tel: 02477- 429712/13/14. Fax: 02477-433045
E-mail: Inevh@satlink.com o vacunafha@yahoo

Firmantes ARGUMEDO, ALCIRA SUSANA - DONDA PEREZ, VICTORIA ANALIA CARDELLI, JORGE JUSTO - LOZANO, CLAUDIO - ITURRASPE, NORA GRACIELA - MERCHAN, PAULA CECILIA - BENAS, VERONICA CLAUDIA - VIALE, LISANDRO ALFREDO - SOLANAS, FERNANDO EZEQUIEL - MACALUSE, EDUARDO GABRIEL - LINARES, MARIA VIRGINIA.

Giro a Comisiones ACCION SOCIAL Y SALUD PUBLICA.

La Cámara de Diputados de la Nación

RESUELVE

Solicitar al Poder Ejecutivo que a través del organismo que corresponda, informe sobre los siguientes puntos:

1. De las vacunas incluidas en el calendario del Plan Nacional de Vacunación que se aplican en nuestro país, cuáles de ellas son de producción nacional y cuales son importadas.

2. Cuál es el monto de la erogación destinada a la importación de vacunas en los últimos cinco años.

3. Con que técnicas se producen las vacunas importadas.

4. Cuáles son las estrategias existentes respecto a la provisión y distribución de vacunas, tanto las incluidas en el Programa Nacional de Vacunación, como así también sobre las que no están incluidas en el mismo.

5. Dado el cambio a nivel mundial del mercado de vacunas, informe si se está planteando alguna alternativa de producción local.

6. Qué tipo de vacunas están en proceso de investigación y desarrollo en los laboratorios de producción pública estatal y cual es el presupuesto previsto para impulsar este desarrollo.

7. Qué tipo de vacunas se están produciendo a nivel nacional en laboratorios de producción pública estatal y que cobertura poblacional brindan.

FUNDAMENTOS

Señor presidente:
La producción de vacunas forma parte de la industria farmacéutica con particularidades propias, lo que se explica teniendo en cuenta que en el año 2002, solamente el 1,5% del consumo mundial de medicamentos correspondió a vacunas.

La necesidad de controlar epidemias de enfermedades infecciosas que castigaron distintos puntos del planeta durante el siglo XX, muchas veces vinculadas a grandes masas de poblaciones pobres, dio un gran impulso a la producción y difusión de vacunas. Existió una demanda para este producto que si bien exigía calidad, también requeria grandes magnitudes y precios bajos. De este modo se desarrollaron las técnicas hoy denominadas tradicionales y que en la actualidad son muy difundidas. En este período la producción de vacunas en general estuvo enmarcada dentro de las llamadas políticas públicas de salud.

La dinámica del mercado de vacunas fue estable hasta fines de la década del 70 del siglo pasado; desde esta época la producción de vacunas, de modo similar a la industria farmacéutica, ha sido atravesada por

un cambio de paradigma tecnológico que la modifica profundamente, en tanto el advenimiento en esos años de la biología molecular como disciplina emergente junto con las técnicas de ingeniería genética, son determinantes de su estado actual.

Esta industria ha tenido un crecimiento sin precedentes en la primer década del nuevo siglo, la investigación y desarrollo de las grandes empresas farmacéuticas ha sido la responsable principal de materializar el conocimiento en productos concretos y como consecuencia de la expansión de este mercado. (Organización Mundial de la Salud, 2009).

En la actualidad existe una gama de técnicas con las que se producen vacunas, que responden a distintos momentos del desarrollo tecnológico

a) Técnicas clásicas o vacunas tradicionales:

1. Vacunas atenuadas vivas

2. Vacunas inactivadas

3. Toxoides

b) Técnicas de avanzada

El avance tecnológico aplicado en esta industria permite superar una serie de limitaciones técnicas del modelo tradicional. Recientemente, con la tecnología de ADN recombinante se trabaja en una nueva generación de vacunas que supera a las anteriores. Estos avances permiten a los investigadores contemplar nuevas estrategias en el desarrollo de vacunas logrando menores costos y mayor seguridad (hepatitis B) o reconstruir una nueva partícula viral segura (HPV).

La participación de organismos internacionales ha tenido gran relevancia en la conformación actual de este mercado. La Organización Mundial de la Salud, así como el Fondo de Naciones Unidas para la Niñez (Unicef) y la Organización Panamericana de la Salud (OPS), están embarcadas en el objetivo de generalizar la inmunización de la población infantil a escala mundial. Existen evidencias de estas acciones en los importantes niveles de inmunización logrados, de acuerdo al informe WHO de 1996 el 80% de la población infantil mundial está vacunada contra la difteria, tétanos y pertussis (Triple bacteriana DTP) y también contra sarampión y poliomielitis se verifica una creciente utilización de las vacunas HB y Hib. Los esfuerzos de estos organismos se han plasmado en una serie de iniciativas de modo de contribuir con la atención de la demanda de vacunas de países subdesarrollados o emergentes. Con este objetivo realizan grandes compras de vacunas lo que les permite tener influencia sobre los precios a los que se comercializan. (Temporao, 2002) Estos hechos brevemente detallados indican que actualmente el mercado de vacunas se estructura en base a dos segmentos: las vacunas tradicionales de uso difundido y bajo precio y las vacunas modernas, con precios elevados y de uso restringido. En cuanto a la tipología de las empresas se puede clasificar en tres grupos (Temporao, 2004)

- Empresas multinacionales con alta tasa de innovación tecnológica y amplio acceso a los mercados internacionales.

- Empresas públicas y privadas tradicionales en países industrializados con gran capacidad de producción de vacunas tradicionales y acceso a la producción de nuevos productos a través de licencias.

- Productores públicos de países en desarrollo con una producción significativa de vacunas tradicionales y

potencial acceso a productos modernos.

La industria de vacunas modernas, puede considerarse una industria biotecnológica (Shao, 2010), caracterizada por un alto umbral de acceso para las empresas, dado que su desarrollo involucra diversos actores desde agencias de gobierno, institutos de investigación y desarrollo y empresas manufactureras.

En el segmento de vacunas modernas existen asimetrías entre la industria de países desarrollados y subdesarrollados. De modo similar a la industria bio-farmacéutica, éstas son producto de los diferentes niveles de desarrollo científico- tecnológico de los países y, como consecuencia del manejo de su conocimiento central, pero además la gran incertidumbre que implica los largos períodos que necesita el proceso de desarrollo de una vacuna y sus costos hundidos, así como el requerimiento de activos complementarios que no siempre controla la industria. Todos estos factores constituyen una profundización de las asimetrías mencionadas.

Cuando se analiza el mercado en términos de volumen, sólo el 14 % de las vacunas requeridas para satisfacer la demanda global proviene del segmento de vacunas modernas, por lo demás existe un gran mercado del órden del 80% de la población mundial, satisfecho por productores de vacunas tradicionales de bajo costo.

El informe de la Organización Mundial de la Salud del 2009 da cuenta de un crecimiento de la participación en el mercado de vacunas tradicionales de productores de vacunas de países en desarrollo. Al respecto menciona que, en el año 2000 el 39% de las dosis de vacunas compradas por organismos internacionales provino desde oferentes de esos países y en el año 2007 esta proporción fue del 60%.

Los países emergentes se ajustan a "planes nacionales de vacunación" de los que forman parte vacunas contra enfermedades infecciosas, altamente contagiosas y para las que existen vacunas tradicionales hace ya muchos años. Estos planes son atendidos por los gobiernos con la participación de los organismos internacionales antes mencionados. Además existen vacunas para enfermedades no incluídas en los programas nacionales que participan de una inmunización no planeada, son distribuídas por los canales tradicionales de las empresas y compradas en forma directa por la población.

A nivel global el mercado de vacunas modernas ha seguido claramente el comportamiento oligopólico de la "gran industria farmacéutica", dónde algunas de sus empresas son grandes jugadores en este segmento. En este contexto esa "gran industria" ha dejado de lado la producción de vacunas económicas de uso difundido, reservando sus antígenos para vacunas más complejas o modernas.

Se abriría de este modo una oportunidad de producción de antígenos al mundo subdesarrollado, mostrando ventanas de oportunidad para países de desarrollo científico tecnológico intermedio que se propongan como objetivo participar de este mercado. Dentro de los objetivos de este trabajo se plantea conocer en detalle la situación en que se encuentra la provisión de vacunas en la Argentina. La realidad local es parte de la realidad global donde los dos segmentos de vacunas participan. Según el criterio adoptado por el Ministerio de Salud de la Nación las vacunas que conforman el "Calendario Nacional de Vacunación" en nuestro país son, a partir del segundo semestre del año 2009:

- BCG (tuberculosis)

- HB (Hepatitis B)

- DPT-HB-Hib: (Pentavalente) Difteria, tétanos, pertussis, Hep B, Haemophilus influenza b

- DPT-Hib: (Cuádruple) difteria, tétanos, pertussis, Haemophilus influenza b).

- OPV (Sabin) Vacuna antipoliomelitica oral.

- SRP: (Triple Viral): Sarampión, Rubéola, Parotiditis

- HA (Hepatitis A)

- DPT: (Triple Bacteriana): difteria, tétanos, pertussis.

- Dtap: (Triple bacteriana acelular)

- DT: (Doble bacteriana) Difteria y tétanos

- SR: (Doble viral) Sarampión, Rubéola

- FA: (Fiebre Amarilla) Una dosis para residentes a zonas de riesgos.

- FHA: (Fiebre Hemorrágica Argentina) una dosis para residentes o viajeros a zonas de riesgo.

La provisión de estas vacunas es responsabilidad del Estado así como asegurar la vacunación en todo el país.

Además existe la distribución de vacunas no planeadas o no incluídas en el Programa Nacional de Vacunación cuya provisión se realiza generalmente a través de las cadenas de distribución farmacéuticas.

En el país existen solamente dos organismos públicos que producen vacunas: el Instituto Biológico de La Plata, dependiente del Ministerio de Salud de la Provincia de Buenos Aires y el Anlis- Malbrán. Estos institutos cuentan con años de experiencia, conocimiento y aprobación de normas técnicas en la fabricación de algunas vacunas de acuerdo a normas internacionales, que son monitoreadas por la OMS y la OPS. Las técnicas con las que se producen vacunas localmente son clásicas o tradicionales. De todas formas son muy pocas las vacunas y la cantidad de dosis que producen estos organismos; el resto son importadas de países de desarrollo científico intermedio como India, Bulgaria y Dinamarca, entre otros.

A modo de síntesis puede decirse que los grandes jugadores que controlan el segmento de vacunas modernas han construido barreras para el acceso al conocimiento científico tecnológico, siendo ésta una de las barreras más relevantes, además de los controles de las redes de distribución y de los gastos en publicidad y propaganda. Esto configura un mercado con altos márgenes de beneficios y con un uso limitado para los países desarrollados y algunos de desarrollo intermedio.

No obstante, más del 80% de la población mundial es abastecida con vacunas tradicionales; de este modo, cabe interrogarse si la entrada de las empresas de países en desarrollo en este segmento de vacunas modernas, es complementaria al desarrollo en el segmento tradicional.

De existir complementariedad, el segmento tradicional sería el punto de partida de una estrategia dual para países de desarrollo científico- tecnológico intermedio que se propongan desarrollar vacunas biotecnológicas y de este modo generar una plataforma de lanzamiento a esta industria.

Por todo lo expuesto solicito a mis pares el acompañamiento del presente proyecto.

Bibliografía

Temporão, J (2002). O Programa Nacional de Imunizações (PNI): Origens e desenvolvimiento. História, Ciência, Saúde. Manguinhos, Rio de Janeiro, vol 10, (suplemento 2), p 601-17.

Temporão, J (2004). ENSP/FIOCRUZ. I Seminário sobre o complexo industrial da Saúde/BNDES. A Indústria de Vacinas. O Brasil no Contexto Mundial

WHO, UNICEF; World Bank. 2009. State of the world's vaccines and immunization Third edition. Geneva. http://whqlibdoc.who.int/publications/2009/9789241563864_eng.pdf -

Zhao, M (2010). Study on Investment Opportunities and Risks in Vaccine Industry. Investment Banking Division of Zero2IPO Group. En http://www.zero2ipo.com.cn/en/n/2010-1-8/20101893747.shtml

[3] Joklik, Wolfgang y otros. Microbiologia-Zinsser. Bs As, Médica Panamericana, 1987, p 1233-6.

[4] Tassi, Virginia y otros, "Fiebre y Exantemas", publicación digital (2007) de la 1° Cátedra de Clínica Médica y Terapéutica y de la Carrera de Postgrado de especialización en Clínica Médica. Facultad de Ciencias Médicas. Universidad Nacional de Rosario. Ver: www.clinica-unr.org.

[5] Elsner, Boris y otros. Lecciones de Patologia. Bs As, López Libreros Editores, 1986, p 720-1.

[6] Joklik, W y otros. Microbiologia-Zinsser ...

[7] Dijo Leónidas A Barrera Oro: "señalo el 'temblor de lengua", asociado con los demás síntomas, como el signo patognomónico de la virosis hemorragica argentina o mal de los rastrojos".

[8] Agnese, Graciela, "Una rara enfermedad alarma a la modesta población de O'Higgins. Análisis del discurso de la prensa escrita sobre la epidemia de Fiebre Hemorrágica Argentina de 1958", en publicación online: Revista de Historia & Humanidades Medicas. Bs As, vol 3, n° 1, julio de 2007. Ampliar con, de la misma autora, "Primeros investigadores de la FHA, 1953-63", en: Duodécimo Congreso Nacional y Regional de Historia Argentina (La Plata, 21-3 de agosto de 2003). Bs As, ANH, 2003 (separata).

[9] Arribalzaga (1913-1985), denominó al mal: "gripe maligna".

[10] Véase el relato de: Muhlmann, Miguel M, "Veinticinco años de la primera denuncia del mal de O'Higgins. FHA. Su historia", en: Boletin de la Academia Nacional de Medicina. Bs As, vol 61, 1º semestre 1983, p 205-22. Dice el autor: "Hace 25 años -el 5 de junio de 1958- (denuncié) ... por primera vez en el país, a través del diario 'La Razón', una nueva entidad patológica ... mal de O'Higgins, conocida como FHA".

[11] Datos aportados por Leónidas A Barrera Oro.

[12] Datos aportados por Gabriel F Barrera Oro.

[13] Comunicaciones orales de Leónidas A Barrera Oro.

[14] Barrera Oro, Julio Guido. Escritos presentados a una revista de divulgación general, Leoplan, en una entrevista en 1959. Su relato alude al año 1958. (No disponemos de más datos).

[15] Barrera Oro, Julio Guido. Escritos presentados a una revista ...

[16] Datos aportados por Leónidas A Barrera Oro, según papeles de Julio G Barrera Oro.
Nota: las tucuras son insectos, cercanos a las langostas.

[17] Datos aportados por Leónidas A Barrera Oro, según papeles de Julio G Barrera Oro.

[18] De los papeles de Julio G Barrera Oro.

[19] Comentarios de Leónidas A Barrera Oro.

[20] Estimamos que se refiere al Dr Pedro D Martini, miembro de la comisión que falleció infectado.

[21] Comentarios de Leónidas A Barrera Oro.

[22] Parodi (1909-1969).

[23] Martone, Francisco José. 50 años de sanidad argentina (vista desde el Congreso de la Nación). Bs As, El Ateneo, 1989 (preferentemente p 104-6).

[24] Maiztegui (1931-1993), en 1990 recibió el premio Cediquifa, en Farmacología, instituido en conmemoración del Día del Investigador: 10 de abril, natalicio de Bernardo A Houssay. En 1993, el premio internacional Dr Sabino Di Rienzo (amplíese con lo dicho en la cita nº 32).

[25] Cfr: Agnese, Graciela, "Entre controversias científico-médicas y movilizaciones populares. Población epidémica y vacunas contra la FHA", en: Actas de las XIXª Jornadas de Historia de la Medicina (realizadas en Bs As, octubre de 2009, responsabilidad del Instituto, Cátedra y Ateneo de Historia de la Medicina. Departamento de Humanidades Médicas). 2010, p 21-31.

[26] Le puso ese nombre en honor a su hermano Leónidas A, quien tanto colaboró en el proceso de su enfermedad y la convalecencia. Tiene que ver con el apodo con el que se lo conocía.

[27] Feuillade, M R y D A Enria, "Análisis de la utilidad de la vacuna Candid 1 en la prevención de la Fiebre Hemorrágica Argentina en niños", en: Revista Panamericana de Salud Pública, 2005, vol 18, nº 2, p 100-6.

[28] Feuillade, M R y D A Enria, "Análisis de la utilidad de la vacuna ...".

[29] (Diario) La Nación. Bs As, 19 de octubre de 1987. Nota de Jacobo Brailovsky.

[30] (Diario) La Nación. Bs As, 29 de noviembre de 1990, p 5 y 26.

[31] (Diario) La Nación. Bs As, 30 de noviembre de 1990. Nota de Fernán Saguier.

[32] (Diario) Ámbito Financiero. Bs As, 5 de julio de 1993. Nota titulada: "Premio Internacional al doctor Julio Maiztegui. El galardón fue entregado por el ministro Alberto Mazza" (en representación del presidente Carlos S Menem).

[33] (Diario) La Nación. Bs As, 28 de octubre de 1996. Nota de Jorge O Manchiola.

[34] Ver en el Anexo el Proyecto de Resolución (para la producción de la primera vacuna argentina contra la FHA, Candid 1; H Cámara de Diputados de la Nación; 29 de agosto de 2006).

[35] (RA) Ministerio de Salud de la Nación. Boletín Epidemiológico Periódico. Bs As, nº 8, abril de 2004.

[36] Decreto-ley nº 3283 (marzo de 1957) y modificado por el nº 16.145 (diciembre 1957).

[37] Pirosky, Ignacio. 1957-1962. Progreso y destrucción del Instituto Nacional de Microbiología. Bs As, Eudeba, 1986, p 194.

[38] Pirosky, Ignacio. 1957-1962. Progreso y ...., p 196-7.

[39] Decreto nº 3577, del 23 de abril de 1962, firmado por el presidente José Maria Guido y el ministro Tiburcio Padilla; el interventor fue Alfredo Benno Fischer. Poco después otros fueron cesanteados o renunciaron (como el caso el matrimonio Celia y César Milstein).

[40] Pablo Kreimer le llama a este lapso: El vacio. Completar con la obra de este autor (Ciencia y periferia. Nacimiento, muerte y resurrección de la biología molecular en la Argentina. Bs As, Eudeba, 2010, preferentemente p 86 y ss).

[41] Completar con: Agüero, Abel L, Alfredo G Kohn Loncarica, Norma Isabel Sánchez y José M Trujillo "Contribuciones originales de la medicina argentina a la medicina universal", en: publicación online, Revista de Historia & Humanidades Médicas, de la Cátedra e Instituto de Historia de la Medicina. Departamento de Humanidades Médicas. Bs As, 3° época, vol 3, nº 1, julio 2007. FM, UBA, www.fmv-uba.org.ar/histomedicina.

LA APLICACIÓN DE LA TEORÍA QSAR/QSPR EN LA PREDICCIÓN DE ACTIVIDADES BIOLÓGICAS Y PROPIEDADES FISICOQUÍMICAS.

I- INTRODUCCIÓN Y PROPÒSITOS GENERALES

Rafael Villamayor, Pablo R. Duchowicz, Eduardo A. Castro*

INIFTA, División Química Teòrica, Suc.4, C.C. 16, La Plata 1900, Buenos Aires, Argentina
* Autor Correspondiente (eacast@gmail.com)

RESUMEN
En esta serie de trabajos se propone ofrecer una descripción panorámica acerca de los actuales empleos de la Teoría de las Relaciones Cuantitativas de Estructra Actividad (Propiedad) (QSAR/QSPR) para predecir actividades biológicas y propiedades fisicoquímicas de las sustancias químicas. En esta primera parte se brinda una introducción general al tema, señalando las principales características de esta metodología así como sus campos de aplicación.

ABSTRACT
In this series of articles we give an overview on the present applications of the Quantitative Structure Activity (Property) Relationships (QSAR/QSPR) to predict biological activities and physical chemistry properties of chemical substances. In this first part we offer a general introduction to this theme, pointing out the main features of this methodology as well as its application fields.

1. La Teoría QSAR/QSPR

El continuo interés por lograr predecir las distintas propiedades fisicoquímicas, biológicas y farmacológicas en sistemas reales conduce indudablemente a la aplicación de métodos derivados de la Mecánica Cuántica, con el fin de representar adecuadamente el fenómeno involucrado. Esto se traduce en la necesidad de tener en cuenta todas las interacciones presentes en el sistema físico de partículas, lo cual hoy por hoy parece ser una tarea harto dificultosa, en vista de que los cálculos mecano-cuánticos actuales sólo pueden resolverse con buena aproximación cuando el sistema involucra unos pocos átomos no-interactuantes. Si bien el uso de aproximaciones matemáticas permite resolver el problema de partículas interrelacionadas entre sí, debido a la incertidumbre de dicho método, no siempre será posible justificar la calidad de los resultados encontrados. Por otro lado, los cálculos mecanocuánticos ayudan a la comprensión de los aspectos mecanísticos que originan a las propiedades en cuestión, pero no resultan la herramienta adecuada para el estudio del efecto que tiene la estructura molecular sobre las propiedades macroscópicas de las sustancias

químicas.

La Teoría QSAR/QSPR, (Relaciones Cuantitativas Estructura-Actividad/Estructura-Propiedad) ofrece una alternativa a la hora de calcular las propiedades de una colección de moléculas. Cuando nos detenemos a observar un conjunto de estructuras moleculares junto con sus propiedades experimentales medidas, la pregunta inmediata que surge es ¿existirá una correlación directa entre la propiedad y la estructura de estas sustancias? La respuesta es afirmativa, y es la hipótesis principal de la Teoría QSAR/QSPR. La misma es una hipótesis matemática fundamentada en el hecho de que la estructura de una molécula es la principal responsable de sus propiedades químicas, fisicoquímicas, biológicas o farmacológicas[1-3]. Quizás una de las premisas fundamentales de la teoría es el Principio de Similaridad Estructural, que establece que estructuras moleculares similares poseen propiedades similares, mientras que estructuras moleculares diferentes manifiestan propiedades diferentes[4]. Si bien es conocido desde hace mucho tiempo el hecho de que distintas sustancias tienen diferentes efectos biológicos, el avance en la determinación de estructuras permitió establecer relaciones estructura-actividad (SAR), las cuales evidencian ciertos efectos en las actividades biológicas a partir del cambio en la estructura química de un determinado compuesto.

Los modelos QSAR también nacen en el campo de la Toxicología. De hecho, los intentos por cuantificar relaciones entre la estructura química y la toxicidad aguda han sido parte de la literatura toxicológica por más de 100 años. Las primeras evidencias se remontan al año 1863, cuando en la defensa de su tesis en la University of Strasbourg, Strasbourg, Francia, J. Cross notó las relaciones existentes entre la toxicidad de alcoholes alifáticos primarios y su solubilidad en agua. Esta relación demuestra el axioma central del modelado de la relación estructura-toxicidad. Por lo tanto, existen interrelaciones entre estructura, propiedades y toxicidad. Casi un siglo después Corwin Hansch et al[5], publicó su famoso artículo sobre la actividad biológica de grupos de compuestos congéneres y con ello sentó la base para el desarrollo de la actual Teoría QSAR/QSPR.

La Teoría QSAR/QSPR busca cuantificar las relaciones SAR a través del desarrollo de modelos, y combina métodos de la Estadística Matemática con la Química Computacional. Tales modelos se vuelven vitales a la hora de predecir el valor de la propiedad de una sustancia si ésta es desconocida por resultar difícil de adquirir, sea por su inestabilidad química, su toxicidad respecto de la salud humana, su costo económico, etc. Así también la teoría ha sido ampliamente utilizada para el diseño y optimización de compuestos tipo-droga, y hasta fue fructíferamente empleada para inferir resultados sobre los mecanismos de reacción de compuestos orgánicos.

Con el fin de establecer un modelo de cuantificación apropiado, un requisito indispensable es disponer de un conjunto de moléculas para las cuales se conocen perfectamente los valores experimentales de la propiedad estudiada. El diseño de un modelo implica su calibración y su posterior validación. La calibración establece con exactitud la correspondencia entre la estructura y la propiedad analizada a través de la creación del modelo y determinación de los parámetros ajustables de los que depende. La función matemática (lineal o no-lineal) que cuantifica la relación estructura-propiedad se elige de forma arbitraria y la simplificación del modelo matemático dependerá de aquella expresión que determine las mejores predicciones. La validación certifica la veracidad del modelo obtenido, es decir, verifica si posee o no un poder predictivo sobre moléculas no contempladas en el ajuste del modelo, y que también deben poseer información conocida de la propiedad experimental.

Pero, ¿cómo representar fielmente las relaciones entre la estructura y la propiedad? Desafortunadamente no existe una vinculación directa entre ambas características, por lo cual la teoría se vale de distintos índices numéricos que codifican la información estructural y ayudan a establecer las relaciones buscadas, estos índices son los denominados descriptores moleculares.

2. Los Descriptores Moleculares

Más estrictamente, un descriptor molecular es el resultado final de una lógica y de un procedimiento matemático que transforma la información química codificada dentro de una representación simbólica de una molécula en un número útil o el resultado de algún experimento estandarizado[6]. Estas variables pueden ser teóricas o experimentales, pueden describir a la molécula como un todo (descriptores globales) o sólo pueden representar un fragmento presente en ella (descriptores fragmentos). Generalmente, un gran número de descriptores moleculares surgen de diferentes teorías, tales como la Teoría de Orbitales Moleculares, la Teoría de Grafos y la Mecánica Cuántica, entre otras.

Ahora bien, puede suceder que una combinación apropiada de números describa adecuadamente la propiedad en cuestión, pero que no dejen de ser eso, sólo "simples números". Así, es requisito fundamental que los descriptores posean algún tipo de interpretación química, y si ese no fuera el caso, que sí puedan derivarse en base a la estructura. Un ejemplo clásico de descriptor lo constituye el número de átomos de una especie química en la molécula, como la cantidad de átomos de carbono en una familia de bencenos o él numero de átomos de cloro en especies clorofluorocarbonadas; la cantidad de enlaces C-C puede ser otro ejemplo de descriptor. Otros descriptores relacionados con propiedades fisicoquímicas pueden ser el índice de refracción, las entalpías de vaporización (ΔHv), el coeficiente de partición octanol/agua (Kow), los puntos de ebullición, los volúmenes molares, etc.

A continuación describiremos los rasgos más relevantes de algunas de las familias de descriptores moleculares más frecuentemente utilizados en la representación de la estructura molecular. No profundizaremos en detalle en cada una de ellas debido a lo amplio y extenso del tema. El lector interesado en este asunto podrá consultar la bibliografía señalada en este análisis.

2.1. Descriptores de la Teoría de Grafos Química

La Teoría de Grafos es una rama de la Matemática Discreta relacionada a la topología y a la combinatoria, y está vinculada con la manera en que los objetos están conectados. Un grafo es una representación bidimensional de la molécula. Estructuralmente, un grafo puede verse como un conjunto de vértices o nodos, unidos por medio de aristas o arcos, en la representación molecular los nodos serían los átomos y las aristas los enlaces. Por ejemplo en el benceno los átomos C son los nodos y los enlaces C-C las aristas.

Los descriptores que se obtienen a partir de la Teoría de Grafos sólo proporcionan información de constitución y conectividad y, por tanto, no pueden discernir isómeros de una misma molécula. Se pueden definir diversos tipos de índices topológicos y entre los más conocidos tenemos a los dos siguientes:

Indice de Wiener (W)[7]

$W = \frac{1}{2}\sum_{ij} d_{ij}$ donde ${}_{ij}$ representa la distancia topológica entre los vértices v_i y v_j, si se considera el camino de longitud más corta. La longitud u orden del camino es el número de aristas que lo componen.

Indice de conectividad molecular (χ)[8]

$\chi = \frac{1}{2}\sum_{ij}\left(\deg_i \deg_j\right)^{-0.5}$ donde $\deg_i$ es el grado de degeneración del vértice v_i y representa el número de vértices adyacentes al mismo.

2.2. Indices de la Teoría de la Información

A menudo sucede que gran cantidad de los índices topológicos calculados poseen alto grado de degeneración. El concepto de degeneración de un descriptor molecular se aplica a aquellos descriptores que posean el mismo valor numérico para estructuras diferentes. La Teoría de la Información ofrece una alternativa para disminuir el grado de degeneración de los descriptores topológicos. La aplicación se basa en darle a la molécula representada por un grafo una cierta distribución de probabilidad respecto a la complejidad que posea, y desde allí aplicar la Teoría de la Información.

2.3. Descriptores para Interacciones Químicas

Estos descriptores caracterizan las interacciones químicas que participan en la molécula tanto a nivel global como local, es decir, refiriéndose a un sector de la molécula o tratándola como un todo. Estas interacciones implican cambios topológicos, geométricos y electrónicos, por lo cual los descriptores suelen combinar algunos de estos aspectos.

2.4. Descriptores del programa Dragon[9]

El programa Dragon® ofrece la posibilidad de calcular un gran número de descriptores moleculares agrupados en diferentes familias. A su vez, la lista de descriptores proporcionados puede ser organizada como cerodimensionales (0D), unidimensionales (1D), bidimensionales (2D), y tridimensionales (3D). Nosotros utilizaremos esta última clasificación para simplificar la descripción. Los descriptores calculados en este trabajo son obtenidos con la aplicación de este programa y son cantidades teórico-definidas y podemos destacar que no se utilizan descriptores experimentales.

Descriptores 0D: describen solamente la constitución de la molécula, pero no dicen nada sobre la conformación ni tipo de conectividad presente. Los más simples son el número de átomos de un determinado tipo, el número de enlaces y el peso molecular, entre otros.

Descriptores 1D: describen fragmentos de las moléculas constituidos por el agrupamiento de sus átomos constituyentes.

Descriptores 2D: utilizan una función de autocorrelación bidimensional que contiene la topología del grafo, y además representa la distribución de una propiedad atómica determinada en la molécula. La propiedad atómica con la que se pesa/pondera al descriptor depende de los átomos presentes en la molécula con lo cual se pueden seleccionar aquellos átomos que dan mayor peso a la variable. Estos descriptores tienen en cuenta las interacciones inter/intra-moleculares.

Descriptores 3D: esta clase tiene en cuenta los aspectos conformacionales de la estructura molecular, considerando de esta manera las propiedades estereoquímicas de las moléculas. Para su cálculo se utilizan estructuras moleculares previamente optimizadas con métodos convenientes, tales como el Método de Campos de Fuerza de la Mecánica Molecular MM$^+$, en combinación con métodos derivados de la Mecánica Cuántica, sean *ab initio* o Métodos de la Teoría de Orbitales Moleculares Semiempírica. Entre estos descriptores citamos las cargas atómicas, la energía del orbital molecular más alto ocupado (ε_{HOMO}) y la energía del orbital molecular más bajo desocupado (ε_{LUMO}), entre otros.

Un descriptor debe cumplir con un conjunto de características tales como:

i. Cálculo sencillo
ii. Invarianza respecto de la traslación y la rotación
iii. Invarianza respecto a la numeración de los átomos
iv. Buena correlación con la propiedad estudiada
v. Bajo grado de correlación con otros descriptores

3. Sobre el Diseño del Modelo

Durante el diseño de los modelos QSAR/QSPR resulta de fundamental importancia seleccionar los descriptores moleculares más influyentes para predecir la propiedad analizada. Existen dos métodos generales para la selección de descriptores moleculares. El primero de ellos consiste en valerse de la experiencia, de las características observables y perceptibles de las moléculas de estudio y del posible mecanismo subyacente. Por ejemplo, la fotohidrólisis es una de las vías principales para la fotólisis de compuestos aromáticos hidrogenados y así varios descriptores químicocuánticos que caracterizan los enlaces C-X fueron calculados y empleados para el desarrollo de modelos QSAR que describan los rendimientos cuánticos de fotólisis de compuestos halogenados[10, 11]. Por otro lado, el segundo método se basa en realizar un estudio combinatorial de los descriptores estructurales y seleccionar aquellos que sean más predictivos.

La ortogonalización de los descriptores moleculares busca facilitar el desarrollo de un modelo óptimo, reduciendo así el número de descriptores objeto de análisis y la dimensión del problema matemático a tratar, por la eliminación de la intercorrelación existente entre dichas variables. Sin embargo, se ha demostrado que la calidad estadística obtenida con el uso de variables no-ortogonales no difiere de la hallada con variables ortogonales.

Las moléculas estándares que constituyen el llamado conjunto de calibración servirán como "moléculas objetivo", pues representan moléculas a las cuales las moléculas de validación deberán imitar, copiar, seguir, aproximarse y lo más deseable, superar en calidad predictiva[12]. Es preciso que las moléculas del conjunto de validación posean estructuras congruentes con las del conjunto de calibración, pues ello influirá directamente en la calidad predictiva del modelo. Una determinada selección de moléculas de calibración y de validación en conjuntos moleculares homogéneos/heterogéneos influenciará considerablemente en los resultados finales que se obtengan con posterioridad con los modelos QSAR/QSPR, y el modelo establecido tendrá algún significado estadístico en la medida que se utilicen conjuntos adecuados.

Finalmente, es esperable que un modelo sencillo que presente error de predicción de la propiedad durante la calibración supere el proceso de validación, en comparación de uno que sea más exacto y sin error de calibración, pues este último se ajusta excesivamente o "memoriza" al conjunto de calibración y de esta manera es incapaz de predecir la propiedad en cuestión durante la validación. Se busca entonces que los errores cometidos por el modelo en la etapa de calibración sean similares a los encontrados durante la etapa de su validación.

4. Objetivo Específicos

El objetivo principal de la presente serie de artículos consiste en estudiar diferentes técnicas estadísticas de diseño molecular que permitan el armado de conjuntos moleculares de calibración y validación balanceados, es decir, conjuntos que posean similares errores de predicción de la propiedad. Se busca así seleccionar la metodología que mejor funcione y así implementarla para el trabajo de investigación QSAR/QSPR cotidiano. Para ello, se abordan las técnicas: Análisis de Agrupamiento Jerárquico, Análisis de Componentes Principales, Análisis Discriminante Lineal, Análisis de Agrupamiento k-Medias y k-vecinos más cercanos. La formulación de relaciones estructura-propiedad está basada en la técnica del Análisis de Regresión Lineal, y considera los aspectos multidimensionales de la estructura por medio del análisis de más de mil descriptores moleculares calculados con el programa Dragon. Se compara la bondad de estos métodos clasificadores de objetos sobre tres bases de datos diferentes, a saber: solubilidades acuosas de 166 compuestos orgánicos tipo-droga, 128 actividades anti-SIDA-1 de compuestos orgánicos, y 470 toxicidades acuosas en compuestos alifáticos heterogéneos.

Referencias

1. King, R. B., "Chemical Applications of Topology and Graph Theory". Studies in Physical and Theoretical Chemistry. Elsevier: Amsterdam, 1983.
2. Sexton, W. A., Chemical Constitution and Biological Activity. D. Van Nostrand: New York, 1950.
3. Hansch, C., Fujita, T., A quantitative approach to biochemical structure-activity relationships. Acc. Chem. Res. 1969, (2), 232.
4. A. M. Johnson, G. M. M., Concepts and Applications of Molecular Similarity. John Willey & Sons: New York, 1990.
5. Hansch, C., Fujita, T., A Method for the Correlation of Biological Activity and Chemical Structure. J. Am. Chem. Soc 1964, 86, 1616.
6. Roberto Todeschini, V. C., Handbook of Molecular Descriptors. WILEY-VCH: Univ. Milano-Bicocca, Italy, 2000.
7. Hosoya, H., A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. Bull. Chem. Soc. Jpn 1971, (44), 2332.
8. Randic, M., Characterization of molecular branching. J. Am. Chem. Soc 1975, (97), 6609.
9. Dragon 3.0. http://www.disat.unimib.it/chm.
10. Chen J W, Q. X., Schramm K -W, Kettrup A, Yang F L., Quantitative structure-property relationships (QSPRs) on direct photolysis of PCDDs. Chemosphere 2000, 45, (2), 151-159.
11. Free S M, W. J. M., A mathematical contribution to structure-activity studies. J Med Chem 1964, 7, (4), 395-399.
12. Randic, M., Resolution of Ambiguities in Structure-Property Studies by Use of Orthogonal Descriptorst. J. Chem. If. Comput. Sci 1991, (31), 311-320.
13. Duchowicz, P. R., Fernández, F. M., Castro, E. A., Alternative Algorithm for the Search of an Optimal Set of Descriptors in QSAR-QSPR Studies. MATCH Commun. Math. Comput. Chem. 2006, 55, 179-192.

LA APLICACIÓN DE LA TEORÍA QSAR/QSPR EN LA PREDICCIÓN DE ACTIVIDADES BIOLÓGICAS Y PROPIEDADES FISICOQUÍMICAS.

II- Técnicas de Clasificación

Rafael Villamayor, Pablo R. Duchowicz, Eduardo A. Castro*

INIFTA, División Química Teórica, Suc.4, C.C. 16, La Plata 1900, Buenos Aires, Argentina
* Autor correspondiente (eacast@gmail.com)

RESUMEN
El diseño de conjuntos moleculares balanceados posee gran importance como paso previo al tratamiento racional empleando la teoría QSAR/QSPR. En el segundo trabajo de esta serie describimos de manera general los métodos màs estudiados y aplicados en el desarrollo de la teoría QSAR/QSPR, tales como: el Análisis de Agrupamiento Jerárquico, el Análisis de Componentes Principales, el Análisis Discriminante Lineal, el Análisis de Agrupamiento k-Medias y k-Vecinos Más Cercanos. Si bien no se han explorado totalmente las numerosas técnicas que hoy aparecen en la literatura estándar, tratamos de considerar a las màs importantes de ellas.

ABSTRACT
The design of balanced molecular sets has a great importance as a previous step to the rational treatment employing the QSAR/QSPR Theory. In the second article of this series we describe through a general manner the most usual and widely applied in the development of the QSAR/QSPR Theory, such as the Hierarchical Grouping Analysis, The Principal Component Analysis, the Lineal Discriminant Analysis, the Grouping of Neighboring k-Medias and k-Vicinal. Although we have not analysed all the numerous available actual techniques appearing in the standard literature, we have considered the most important ones.

1. El Problema de la Clasificación

La clasificación es el proceso de división de un conjunto de objetos en grupos mutuamente excluyentes, de manera tal que los miembros de cada grupo se hallen lo más cerca posible el uno al otro, y los de diferentes grupos lo más lejos posible[1]. La cercanía se mide respecto a una determinada variable que forma la predicción. No existe en general una regla que permita definir la mejor aproximación a un problema de clasificación en particular[2], [3], pero la selección adecuada de los descriptores moleculares clasificadores resulta ser esencial en la Teoría QSAR/QSPR.

Una de las estrategias más comunes para el desarrollo de relaciones cuantitativas específicas está basada en la clasificación de moléculas según su funcionalidad química, lo cual resulta sencillo desde el punto de vista práctico. Sin embargo, puede suceder que dos moléculas determinadas tengan los mismos grupos funcionales y distinto valor en sus propiedades macroscópicas, lo que hace este esquema de clasificación un tanto problemático.

Entre las técnicas de clasificación de datos más difundidas en la literatura encontramos al Análisis de Componentes Principales (PCA)[4], Análisis Discriminante (DA)[5], Análisis de Agrupamiento (CA)[6], y otras pertenecientes al campo de las Redes Neuronales (ANN), como pueden ser las Redes Neuronales de Retro-Propagación (BPNN)[7] o los Mapas de Auto-Organización de Kohonen (SOM)[8]. Por su parte, la Teoría de la Lógica Difusa (FLT)[9], [10] representa una herramienta alternativa del área de la Inteligencia Artificial aplicable a problemas de clasificación, y que consigue modelar razonablemente conceptos difusos relacionados a la incerteza o imprecisión. Dentro de dicha teoría se han reportado los algoritmos de Agrupamiento Difuso (FC)[11] y de Partición Difusa Adaptativa (AFP)[12].

El diseño de conjuntos moleculares balanceados posee gran interés como paso previo al tratamiento racional QSAR/QSPR. En el segundo trabajo de esta serie describimos de manera general los métodos más estudiados y aplicados en el desarrollo de la teoría QSAR/QSPR, tales como: Análisis de Agrupamiento Jerárquico, Análisis de Componentes Principales, Análisis Discriminante Lineal, Análisis de Agrupamiento k-Medias y k-Vecinos Más Cercanos. Si bien no se han explorado totalmente las numerosas técnicas que hoy aparecen en la literatura, tratamos de considerar el mayor número de ellas.

2. Análisis de Componentes Principales (PCA)

Uno de los problemas inherentes en Estadística Multivariable es la dificultad de visualización de datos que dependen de gran cantidad de variables. Si bien un simple gráfico en dos o tres dimensiones facilita la interpretación, la existencia de cuatro o más variables dificulta la visualización de las relaciones existentes. Afortunadamente, en un conjunto de datos con muchas variables los grupos de variables a menudo se mueven juntos. Una consecuencia de ello es que más de una variable puede ser la fuerza impulsora que gobierna el comportamiento del sistema, con lo cual estamos frente a un problema de redundancia de información.

Es posible simplificar la dimensión del problema matemático mediante la sustitución de un grupo de variables correlacionadas (X) por una única nueva variable (PC). El Análisis de Componentes Principales es un método cuantitativamente riguroso basado en la correlación de los datos, utilizado para llevar a cabo esta simplificación. El método genera un nuevo conjunto de variables, llamadas componentes principales (PC). Cada uno de los PC es una combinación lineal de las d variables originales, y que resultan ortogonales entre sí, con lo cual no existe información redundante:

$$PC_i = \sum_{j=1}^{d} a_{j,i} X_j \qquad (1)$$

En esta ecuación, PC_i es el i-ésimo componente principal, y a_{ji} es el coeficiente de la *j*-ésima variable para ese componente.

El número de componentes principales coincide con el número de variables independientes utilizadas para derivarlos. En la práctica, para un problema particular se selecciona el número de PC de manera que la suma de las varianzas de los primeros PC exceda el 80% de la varianza total de los datos originales. Cada componente principal se extrae en orden decreciente de varianza explicada por tal componente en el conjunto de datos. Una vez que se ha eliminado la redundancia, sólo los primeros componentes son requeridos

para describir la mayor parte de la información contenida en el conjunto original de datos. Este enfoque ayuda a separar los componentes importantes de aquellos que sólo expliquen una variabilidad al azar.

Existen infinitas formas de construir una base ortogonal para el espacio de los datos, por lo cual: ¿qué hay de especial en el conjunto de componentes principales? Aquí, el primer componente principal (PC1) es un eje sencillo cuya dirección es convenientemente elegida en el espacio. Cuando se proyecta cada observación sobre el eje PC1, los valores resultantes forman una nueva variable denominada coordenada PC1, cuya varianza es máxima respecto de toda posible elección del primer eje. El segundo componente (PC2) es otro eje perpendicular al primero; si se proyectan las observaciones sobre este eje se genera una nueva variable denominada coordenada PC2, cuya varianza es la máxima entre todas las opciones posibles del segundo eje. Los componentes restantes se toman ortogonales a los previamente seleccionados y describen la varianza máxima de los datos.

Ventajas y desventajas del método PCA

Ventajas:

- Puede constituir un camino para determinar la dimensionalidad efectiva de un conjunto de datos.
- Al ser los PC ortogonales entre sí, las comparaciones hechas entre objetos con respecto a sus coordenadas en un dado PC*i* no están correlacionadas con comparaciones que estén basadas en las coordenadas en otro PC*j*.

Desventajas

- Es frecuente no encontrar interpretación alguna de los componentes obtenidos. Cada componente es una combinación lineal de variables que reflejan distintas características de las observaciones.
- Los componentes principales no son invariantes a transformaciones lineales de las variables. Por lo tanto, las componentes se modifican si las variables se estandarizan.

3. Análisis Discriminante Lineal (LDA)

En el Análisis Discriminante, el punto de partida es un conjunto de objetos clasificados en dos o más grupos. De estos objetos, se conocen sus variables atributo. Al reconocer de antemano la existencia de estos grupos, parece lógico pensar que existen variables cuyo valor numérico determina la pertenencia a uno u otro grupo. Los objetivos del Análisis Discriminante son:

i. La identificación de variables atributo que mejor discriminen entre los grupos y la evaluación del poder discriminante de cada una de ellas.

ii. Asignar, con un cierto grado de riesgo, un objeto del que no se conoce su clasificación y del que se conocen las variables atributo.

Como técnica de análisis de dependencia, LDA permite obtener un modelo lineal de causalidad en el cual la variable dependiente puede ser métrica o categórica, y las variables independientes son métricas, continuas y determinan a qué grupo pertenecen los objetos. Se trata de encontrar relaciones lineales entre las variables que mejor discriminen a los grupos iniciales de objetos. Además, se trata de definir una regla de decisión que asigne un nuevo objeto a uno de los grupos prefijados. Para más información sobre el método LDA y la manera con que asignan objetos, ver Apéndice, sección I.

Ventajas y desventajas del método LDA

Ventajas:

- La técnica LDA es fácil de aplicar.

- Las probabilidades de pertenencia a un grupo dado son fáciles de obtener.
- Está disponible en muchos programas estadísticos.

Desventajas:

- Las suposiciones de normalidad e igualdad de varianzas no siempre se cumplen en las variables del modelo.
- La clasificación de nuevas observaciones no es muy eficiente a medida que se incrementa el número de variables del modelo. Se acostumbra a seleccionar variables antes de aplicar LDA.
- Requiere que se especifiquen los grupos del conjunto de entrenamiento del modelo con clases prefijadas.

4. Análisis de Agrupamiento

El análisis de agrupamiento, también llamado análisis de segmentación o análisis de taxonómico, crea grupos o agrupaciones de datos. Estas agrupaciones están formadas de tal manera que los objetos en el mismo grupo son muy similares y los objetos en grupos diferentes son muy distintos. Podemos encontrar distintos tipos de análisis, divididos generalmente en dos grandes categorías:

- Jerárquicos: construyen una jerarquía de agrupamiento

- Particionamiento: el número de grupos se determina de antemano y las observaciones se asignan a tales grupos según su proximidad o cercanía.

4.1. Análisis de Agrupamiento Jerárquico (HCA)

Agrupa los objetos mediante la creación de un árbol jerárquico o dendrograma. El árbol no es simplemente un conjunto de grupos, sino más bien una jerarquía de múltiples niveles, donde los agrupamientos en un nivel dado aparecen unidos como agrupamientos del nivel siguiente. Ello permite decidir el nivel o grado de agrupamiento que resulta más apropiado para la aplicación particular. Uno de los pasos más importantes del HCA lo constituye la búsqueda de similitud o disimilitud entre los objetos en el conjunto de datos, por lo cual existe una gran variedad de formas de calcular esta medida. Para llevar a cabo un HCA se sigue el procedimiento a continuación:

a. Encontrar la similitud o disimilitud entre los objetos

En este paso se calcula la distancia entre cada par de objetos para un método de medida definido. En el caso de un conjunto de datos formado por m objetos, existen $\frac{m(m-1)}{2}$ pares posibles, y las distancias generadas para dichos pares conducen a una matriz distancia o disimilaridad. La medida de distancia entre objetos más comúnmente utilizada es la distancia Euclídea. Sin embargo, uno podría utilizar otras opciones como: distancia Euclídea estandarizada, distancia Mahalanobis, distancia Manhattan, o distancia Minkowski, entre otras, más información en Apéndice, sección II. A veces sucede que en el conjunto de datos utilizados las variables poseen diferentes escalas o diferentes unidades. Estas discrepancias pueden influir directamente a la hora de realizar el cálculo de proximidad, por lo cual como paso previo al cálculo de la matriz distancia es posible estandarizar/normalizar los valores de los datos a fin de utilizar la misma escala proporcional.

b. Agrupar los objetos en el dendrograma

En este paso se enlazan los objetos o agrupamientos más próximos entre sí, mediante una función de enlace

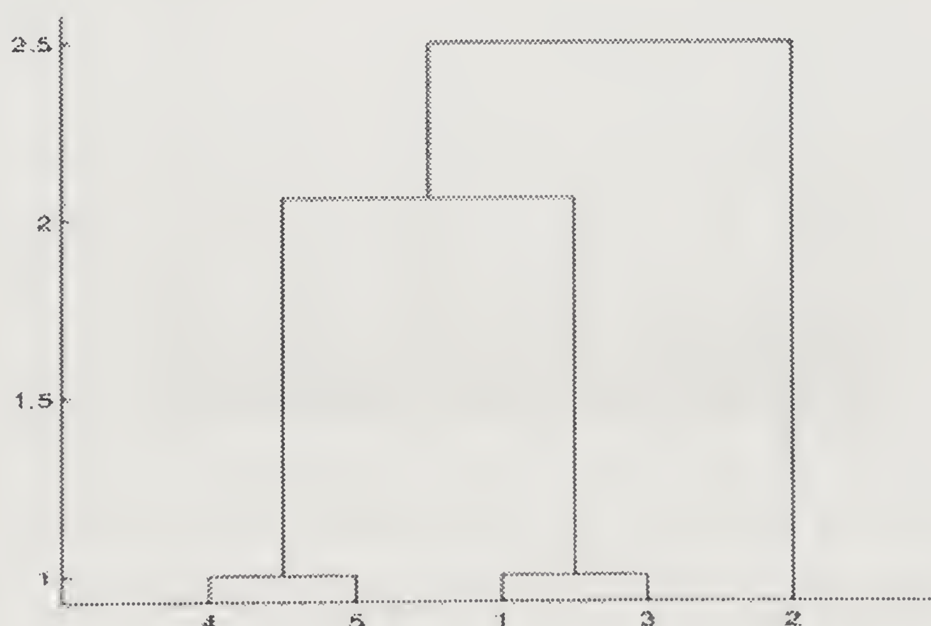
o vinculación. La función de vinculación utiliza la información de las distancias obtenidas en el paso anterior, y asocia inicialmente los pares de objetos más próximos en grupos binarios. A continuación, vincula estos grupos con otros objetos más lejanos para crear agrupamientos binarios de mayor tamaño, hasta que todos los objetos del conjunto de datos original forman el árbol jerárquico. Existen diferentes métodos de vinculación disponibles, los métodos difieren entre sí en la forma de medir la distancia entre agrupamientos. Por ejemplo, el método de Vinculación Individual utiliza la distancia más cercana entre pares de objetos o grupos; otro caso como el método de Vinculación Promedio utiliza la distancia promedio entre todos los elementos en cualquiera de las dos agrupaciones, etc. Más detalles se presentan en el Apéndice, sección III.

c. Especificar el grado de agrupamiento buscado

En general, pueden crearse agrupamientos de datos si se detectan agrupaciones naturales en el árbol jerárquico, o sino a través de realizar un corte horizontal arbitrario del dendrograma. En este último caso, se busca que el corte horizontal intersecte las lineas verticales del gráfico, y esto genera el número de grupos dependiente de la posición del corte.

d.1. Representación gráfica de HCA: dendrograma

La jerarquía creada a través de la generación de agrupamientos binarios mediante las funciones de vinculación puede ser fácilmente entendida cuando se visualiza gráficamente. El dendrograma resultante tiene la siguiente estructura:



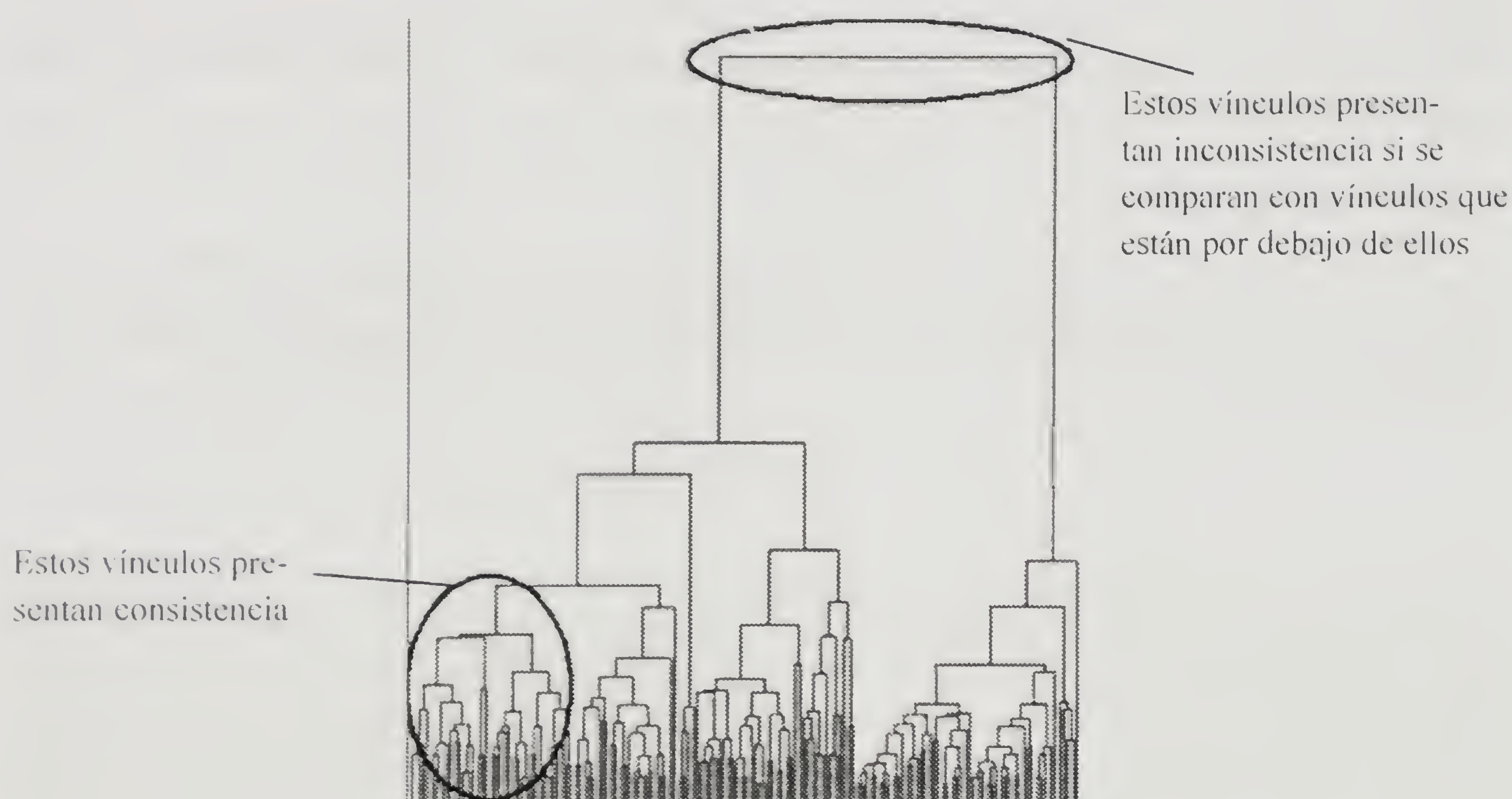
En la figura, los números a lo largo del eje horizontal representan los índices de los objetos en el conjunto de datos original, mientras que el eje vertical mide la distancia. La vinculación entre objetos o grupos se representa como líneas en forma de U invertidas. La altura de U señala la distancia entre grupos.

d.2. Verificación de disimilaridad

En un árbol de agrupamiento jerárquico, cualquier par de objetos en el conjunto de datos original está eventualmente vinculado en algún nivel. La altura de la vinculación en el dendrograma representa la distancia entre ambos objetos, y es conocida como la distancia cofenética para el par de objetos. Una manera de medir la bondad o precisión del árbol de agrupamiento es comparar las distancias cofenéticas con las distancias generadas en la primera etapa del análisis (punto a). Si el agrupamiento es válido, la vinculación de objetos en el árbol jerárquico deberia tener una fuerte correlación con la distancia entre objetos. El coeficiente de correlación cofenético compara ambas distancias, y un valor cercano al valor uno sugiere que la solución de agrupamiento encontrada representa a los datos.

d.3. Verificación de consistencia

Una manera de detectar divisiones naturales en los datos es comparar las alturas de cada enlace en el árbol jerárquico con las alturas de los enlaces vecinos que se encuentran por debajo. Un enlace que está aproximadamente a la misma altura que un enlace que se encuentra por debajo sugiere que no hay divisiones claras entre los objetos unidos. Se dice que estas vinculaciones presentan un alto nivel de consistencia, pues la distancia entre los objetos que se han unido es semejante a la distancia entre los objetos que contienen. Por otro lado, un enlace cuya altura difiere notablemente de la altura de los enlaces inferiores indica que los objetos unidos a ese nivel están mucho más separados entre si de lo que estaban sus componentes cuando se los unió. En el análisis, los vínculos inconsistentes pueden indicar los bordes de una división natural en el conjunto de datos, pues detectan regiones donde la similaridad entre objetos o grupos cambia abruptamente.



El coeficiente de inconsistencia compara la altura de un vínculo en un árbol jerárquico con el promedio de las alturas de los vínculos que están por debajo de él y permite cuantificar la inconsistencia relativa. Los vínculos que unen distintos agrupamientos tienen altos coeficientes de inconsistencia, a diferencia de aquellos que unen grupos indistintos, que tienen un bajo valor del coeficiente.

Ventajas y desventajas del método HCA

Ventajas:

- La medida de la distancia es el único factor que determina al agrupamiento.
- No es un método iterativo, por lo que no se necesitan soluciones iniciales y no existen problemas de optimización que conduzcan a mínimos locales.
- No presenta una sensibilidad apreciable en presencia de ruidos (objetos que posean cierta ambigüedad respecto al agrupamiento al cual pertenecen)[13].

Desventajas:

- Resulta ser un método gráfico, más que analítico. El número de agrupamientos óptimo (k) se obtiene por inspección del gráfico.
- Requiere que se especifiquen los grupos del conjunto de entrenamiento del modelo.

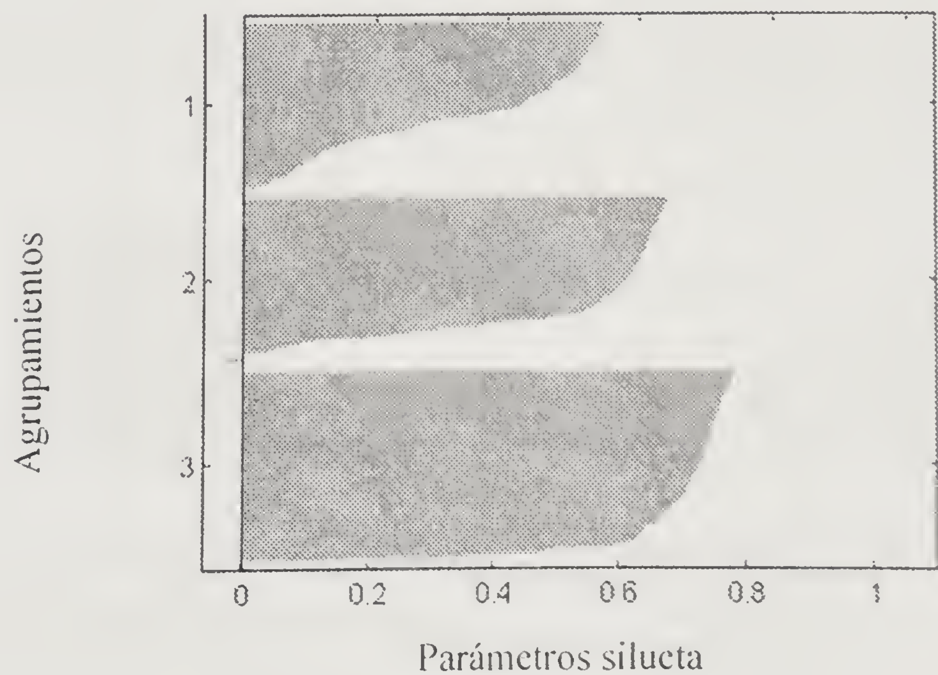
- La interpretación de los resultados de HCA es menos directa. Es necesario armar el dendrograma y recién ahí se podrá saber la clasificación asignada a cada objeto.

4.2. Análisis de Agrupamiento k-Medias

En este método de agrupamiento se particionan los datos de una matriz X en k grupos mutuamente excluyentes. A diferencia de la técnica HCA, el método k-Medias opera sobre observaciones reales en lugar de considerar medidas de disimilitud entre objetos o grupos, y crea por tanto un único nivel de agrupaciones14. Esto hace que k-Medias sea a menudo más conveniente que el agrupamiento jerárquico si se aplica para la clasificación de gran cantidad de datos.

El k-Medias trata cada observación como un objeto que tiene una ubicación en el espacio. Luego de realizar un proceso iterativo, identifica una partición tal que los objetos dentro de cada agrupamiento estén ubicados lo más cerca posible el uno al otro, y también lo más lejos posible a otros grupos. Por supuesto, es posible elegir la medida de la distancia y ello dependerá del tipo de datos que se analicen. Cada agrupamiento de la partición se caracteriza por sus objetos miembros y por su centroide o centro. El centroide para cada agrupamiento es el punto en el que la suma de las distancias de todos los objetos en tal agrupamiento se hace mínima. El tipo de distancia utilizada en forma predeterminada suele ser Euclídea, pero al igual que en HCA es posible escoger distintas opciones. El cálculo de los centroides proporciona diferente resultado dependiente del tipo de medida de distancia empleado. La ubicación final del centroide se determina a través de un proceso iterativo, que generalmente converge a una solución que es mínimo local en la primera etapa del cálculo, pero en la segunda etapa en la mayoría de los casos alcanza un mínimo global, para más información ver Apéndice, sección IV.

Para tener una idea de la calidad de las agrupaciones generadas por k-Medias, se define el valor silueta para cada objeto i ($s(i)$). El valor silueta representa una medida de la similitud que tiene un objeto situado en un grupo dado respecto a otros pertenecientes a grupos vecinos. Su valor numérico cae en el intervalo [-1 1]. Un valor silueta cercano a 1 indica la buena asignación del objeto al agrupamiento, en tanto la disminución del indicador empeora la calidad de la asignación. Si $s(i)$ es cercano a cero, resulta indistinto asignar el punto i a un grupo o a otro grupo vecino. Si $s(i)$ es negativo, sugiere que el objeto se asignó a un grupo erróneo. La definición del parámetro silueta se presenta en el Apéndice, sección V.
A partir de la comparación de la magnitud de los valores silueta, es posible ajustar el valor de k a utilizarse en la clasificación. El siguiente gráfico silueta constituye un ejemplo.



El gráfico representa el número de agrupamientos producidos en función del parámetro silueta.

Se aprecia que muchos de los objetos en el tercer agrupamiento tienen *s*(*i*)>0.6. por lo que dicho grupo está bien resuelto de las agrupaciones vecinas. No obstante. se puede observar que el primer grupo tiene muchos objetos con valores menores de *s*(*i*). y que el segundo grupo presenta inclusive unos pocos valores negativos del parámetro. lo que manifiesta que estos dos agrupamientos no están bien separados.

Ventajas y desventajas del método k-Medias

Ventajas:

- k-Medias puede producir agrupamientos más estrictos que HCA. especialmente si los agrupamientos son de tipo globular15.
- k-Medias es un método más analítico. si se compara con HCA.
- La determinación de los centroides es automática y no se requiere información adicional sobre las clases presentes en el conjunto original de datos.

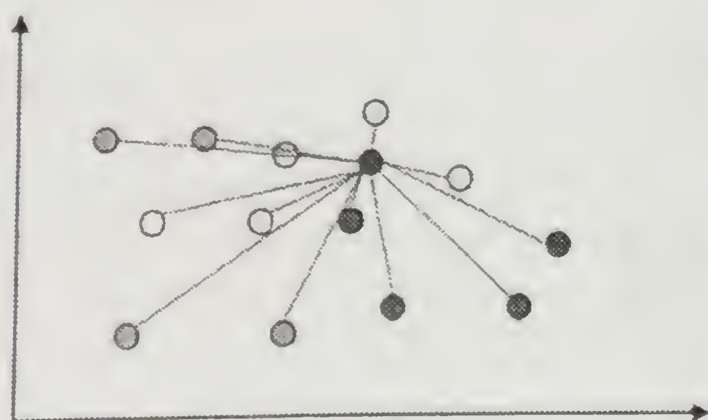
Desventajas:

- Fijar el número de agrupamientos puede hacer dificultosa la selección del valor óptimo de la variable k.
- Alta sensibilidad a la posición inicial de los centroides de las agrupaciones en el método iterativo. Con el fin de obtener una solución óptima. se deben realizar varias pruebas con distintas posiciones iniciales de los centroides.
- No funciona bien con agrupamientos no-globulares.

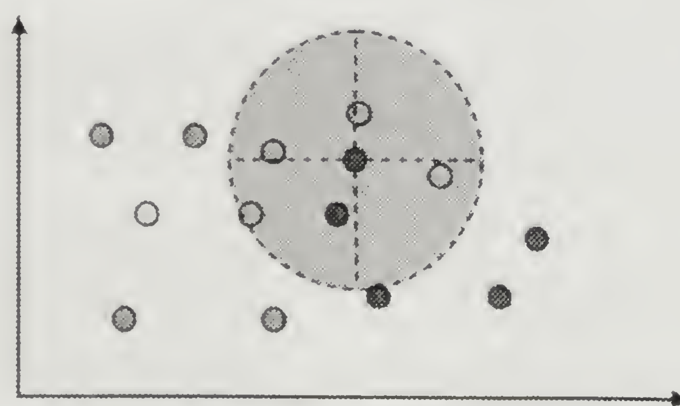
4.3. Análisis k-Vecinos Más Cercanos (k-NN)

El análisis de vecinos más cercanos consiste en estimar el valor de un dato desconocido a partir de las características del dato más próximo. según una medida de similitud o distancia[16]. Este análisis tiene propiedades estadísticas bien establecidas y es fácil de aplicar a sistemas reales[17]. El método de vecinos más cercanos se puede extender si se utiliza no uno. sino un conjunto de datos más cercanos para predecir el valor del nuevo dato. en lo que se conoce como k-vecinos más cercanos.

El k-NN asume que todos los objetos pertenecen a un conjunto de calibración predeterminado. y mediante una medida de distancia elegida se determinan los k objetos más cercanos al objeto que se desea clasificar. Se trata de un algoritmo de aprendizaje inductivo supervisado. en el que se genera una función que asigna las entradas a salidas deseadas. Esto significa que el conjunto de calibración incluye. además de las propiedades multidimensionales utilizadas para el reconocimiento (variables atributo). clasificadores para predecir la clase a la que pertenecen los datos de entrada. Por ejemplo. un objeto es asignado a una determinada clase si ésta es la clase más frecuente entre los k objetos de entrenamiento más cercanos.



Distancias entre el punto a clasificar al conjunto de entrenamiento. El punto negro representa el punto a clasificar.



Generación de un conjunto cuando k=5.

El método k-NN supone que los vecinos más cercanos conducen a la mejor clasificación, esto se hace al considerar todas las variables atributos. El problema de tal suposición es que es posible que existan varios atributos irrelevantes que dominen sobre la clasificación, así los atributos relevantes pierden peso de decisión y la clasificación es incorrecta. Para resolver la cuestión, es posible asignar un peso a las distancias de cada atributo, que transfiere mayor importancia a los atributos más relevantes. Otra posibilidad es tratar de asignar los pesos con objetos conocidos de entrenamiento.

La mejor elección del valor de *k* depende fundamentalmente de los datos; generalmente los valores altos de *k* reducen el efecto de ruido en la clasificación, pero crean límites entre clases parecidas. El valor más adecuado de esta variable corresponde a aquel que provea la mejor clasificación de los datos para la aplicación concreta. La exactitud del algoritmo k-NN puede ser profundamente degradada por la presencia de ruido o características irrelevantes, por lo que todos los datos deben estar apropiadamente estandarizados.

Ventajas y desventajas

Ventajas:

• Simple uso. Como existe un conjunto de entrenamiento con la clasificación de objetos preestablecida en el mismo, la clasificación de nuevos objetos sólo implica la medida de la distancia entre objetos y no requiere de un cálculo iterativo.
• Debido a que se cuenta con más información inicial, la clasificación debería ser más exacta, si las clases iniciales impuestas son las correctas.

Desventajas:

• Requiere información adicional, pues es necesaria la asignación de un conjunto de entrenamiento, para lo cual se debe conocer de antemano las clases iniciales; esta información no siempre está disponible.

Referencias Bibliográficas

1. Mazzatorta, P., Benfenati, E., Lorenzini, P., Vighi, M., QSAR in Ecotoxicity: An Overview of Modern Classification Techniques. J. Chem. Inf. Comput. Sci. 2004, 44, 105.
2. Brown, R. D., Martin, Y. C., Use of Structure−Activity Data To Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. J. Chem. Inf. Comput. Sci. 1996, 39, (3), 572.
3. Marengo, E. T., R, Linear Discriminant Hierarchical Clustering: A modeling and Cross-Validate Divisive Clustering Method. Chemom. Intell. Lab. Sys. 1993, 19, 43.
4. Niemi, G. J., Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology. Kluwer Academic Publishing: Dordrecht, 1990.
5. Hubert, C. J., Applied Discriminant Analysis. Wiley-Intersciene: New York, 1994.
6. L. Kaufman, R., P. J., Finding Groups in Data: An Introduction to Cluster Analysis. Wiley-Interscience: New York, 1990.
7. Hecht-Nielsen, R., Proceedings of the International Joint Conference on Neural Networks. Washington D. C., 1989; p 531.
8. Kohonen, T., Self-Organizing Maps. Springer-Verlag: Berlin, 2001.
9. H.T. Nguyen, W., E. A., A First Course in Fuzzy Logic. Third Edition (2006) ed.; Chapman&Hall/CRC.
10. Zadeh, L. A., Classification and Clustering. Academic Press: New York, 1977.
11. F. Ros, A., K., Pintore, M., Chrétien, J. R., Hybrid Systems for Virtual Screening: Interest of Fuzzy Clustering Applied to Olfaction. SAR&QSAR Environ. Res. 2000, 11, 281.
12. Ros, F., Taboureau, O., Pintore, M., Chrétien, J. R., Development of predictive models by adaptive fuzzy partitioning. Application to compounds active on the central nervous system. Chemom. Intel.

Lab. Syst. 2003, 67, 29.

13. Iye, M., Hopfinger,A. J., Treating Chemical Diversity in QSAR Analysis: Modeling Diverse HIV-1 Integrase Inhibitors Using 4D Fingerprints. J. Chem. Inf. Model. 2007, 47, 1945.

14. Matlab 7.0, The MathWorks, Inc. http://www.mathworks.com

15. Han, J., Kamber, M., Pei, J., In Data Mining: Concepts and Techniques, 2006.

16. Cover, T., Hart, P., Nearest neighbor pattern classification. In IEEE Transactions on Information Theory, 1967; Vol. 13, p 21.

17. D. Aha, D. K., M. Albert, Instance-based learning "Instance-based learning algorithms". In Machine Learning, 1991; p 37.

APÉNDICE

I. Discriminación y Clasificación en LDA

Si consideramos una matriz de atributos o variables independientes **X** para cada objeto que pertenece a una clase G_1 determinada, este conjunto de muestras es el llamado conjunto de entrenamiento o calibración. El problema consiste entonces en hallar una buena predicción de la clase G_1 de un objeto considerado, con el uso de la misma distribución del conjunto de entrenamiento, a través de los valores de las variables X_i .

La obtención de la función discriminante incluye una serie de aproximaciones, a saber:
i. las funciones densidad de probabilidad, $p(\mathbf{X} \mid G_1 = 0)$ y $p(\mathbf{X} \mid G_1 = 1)$, son ambas distribuciones normales. G_1.

ii. si (μ_0, Σ_0) y (μ_1, Σ_1) son los parámetros media y covarianza de G_1=0 y G_1=1, respectivamente, se supone que las covarianzas son iguales.

Función Discriminante Lineal
Es posible encontrar una relación lineal que caracterice a cada objeto según los valores de sus atributos. De esta manera, si tenemos un conjunto de N objetos de los que se conocen D variables explicativas, y se observa que N_1 de ellos pertenecen a la clase C_1, y los N_2 restantes a la clase C_2, con $N_1+N_2=N$, es posible construir una función lineal en base a las D variables y puede usarse para predecir si pertenece a un grupo u otro con una probabilidad determinada. En la función lineal:

$$Z = \sum_{i=0}^{d} \lambda_i X_i \qquad \text{(AI.1)}$$

Z es una variable clasificadora, X_i es la i-ésima variable atributo, y λ_i es su coeficiente. El objetivo principal de tal función lineal desde el punto de vista de la varianza consiste en responder a la pregunta de si dos o más grupos son significativamente distintos uno a otro respecto a la medida de una variable en particular. Debe tenerse presente que si la media de una variable es significativamente diferente en varios grupos, puede decirse que esta variable discrimina bien entre grupos.
En caso de que sea posible identificar más de dos grupos en los datos, pueden estimarse funciones discriminantes múltiples, cada una de ellas similares a la presentada en la Ec. (AI.1). Por ejemplo, cuando se tienen tres grupos, puede estimarse: 1) una función para discriminar entre el grupo 1 y los grupos 2 y 3 combinados; y 2) otra función para discriminar entre el grupo 2 y el grupo 3. Además, se pueden considerar sólo las funciones discriminantes múltiples que resulten más significativas: si se observan los coeficientes es-

tandarizados de las variables de cada una de las funciones escogidas, cuanto mayor sean estos coeficientes, más alta es la contribución a la discriminación especificada. Finalmente, pueden considerarse las medias de las funciones discriminantes significativas para analizar entre cuales grupos éstas discriminan.

II. Tipos de Medida de Distancia

Sea una matriz **X** (mxn) que es tratada como m vectores fila $x_1, x_2, \ldots, x_m$. Varios tipos de medida de distancia que son posibles definir para el par de objetos r y s, con x_r y x_s, se incluyen a continuación:

i. Distancia Euclídea: es la distancia entre dos puntos que se mide en el espacio euclídeo, se define como

$$d_{rs}^2 = (x_r - x_s)(x_r - x_s)' \quad \text{(AII.1)}$$

ii. Distancia Euclídea Estandarizada: cada coordenada en la suma cuadrática se pesa inversamente por la varianza muestral de esa coordenada.

$$d_{rs}^2 = (x_r - x_s)\mathbf{D}^{-1}(x_r - x_s) \qquad \text{(AII.2)}$$

donde **D** es la matriz con elementos diagonales dados por v_j^2, que se refiere a la varianza de la variable X_j sobre los m objetos.

iii. Distancia Mahalanobis: es una forma de determinar la similitud entre dos variables aleatorias multidimensionales. A diferencia de la distancia Euclídea, esta medida tiene en cuenta la correlación de las variables.

$$d_{rs}^2 = (x_r - x_s)\mathbf{V}^{-1}(x_r - x_s) \qquad \text{(AII.3)}$$

donde **V** es la matriz de covarianza muestral.

iv. Distancia Manhattan: aquí la distancia entre dos puntos es la suma de las diferencias (absolutas) de sus coordenadas.

$$d_{rs}^2 = \sum_{j=1}^{n} | x_{rj} - x_{sj} | \qquad \text{(AII.4)}$$

v. Distancia Minkowski

$$d_{rs} = \left\{ \sum_{j=1}^{n} | x_{rj} - x_{sj} |^p \right\}^{\frac{1}{p}} \qquad \text{(AII.5)}$$

En el caso especial p=1, la distancia Minkowski coincide con la distancia Manhattan, y para el caso especial p=2, la distancia Minkowski coincide con la distancia Euclídea.

vi. Distancia Coseno: uno menos del ángulo incluido entre los puntos (en forma de vector).

$$d_{rs} = \left\{ 1 - x_r x_s' \Big/ \left(x_r' x_r\right)^{\frac{1}{2}} \left(x_s' x_s\right)^{\frac{1}{2}} \right\} \qquad \text{(AII.6)}$$

vii. Distancia de Correlación: uno menos la correlación entre los puntos (tratado como secuencias de valores).

$$d_{rs} = 1 - \frac{(x_r - \bar{x}_r)(x_s - \bar{x}_s)'}{\left[(x_r - \bar{x}_r)(x_r - \bar{x}_r)'\right]^{\frac{1}{2}} \left[(x_s - \bar{x}_s)(x_s - \bar{x}_s)'\right]^{\frac{1}{2}}} \qquad \text{(AII.7)}$$

donde $\overline{x}_r = \frac{1}{n}\sum_j x_{rj}$ y $\overline{x}_s = \frac{1}{n}\sum_j x_{sj}$

viii. Distancia Hamming: es el porcentaje de coordenadas que difieren.

$$d_{rs} = \left(\#\left(x_{rj} \neq x_{sj}\right)/n\right) \qquad \text{(AII.8)}$$

III. Métodos de Enlace o Vinculación

En el método HCA, se utiliza una función vinculante que crea un árbol de agrupamiento jerárquico a partir de las distancias entre pares de objetos previamente obtenidas. La función puede utilizar diversos métodos de vinculación, los cuales difieren en la forma que se calculan las distancias entre los agrupamientos.
La solución que se obtiene en HCA es una matriz (m-1)x3 llamada Q, donde m es el número de observaciones en el conjunto original de datos. Las primera y segunda columnas de Q contienen a los índices de los agrupamientos vinculados de a pares, para formar el árbol binario. La tercera columna contiene la distancia de vinculación entre los agrupamientos formados.

La siguiente notación se utiliza para describir los distintos métodos de vinculación:

- Un agrupamiento r es formado a partir de los agrupamientos p y q
- n_r es el número de objetos en el agrupamiento r.
- x_{ir} es el i-ésimo objeto en el agrupamiento r.

a. Vinculación individual, también llamado vecino más cercano, utiliza la menor distancia entre dos objetos en los dos agrupamientos.

$$d(r,s) = \min(dist(x_{ri}, x_{sj}), i \in (i,\dots,n_r), j \in (1,\dots,n_s) \qquad \text{(AIII.1)}$$

b. Vinculación completa, también llamado vecino más lejano, utiliza la mayor distancia entre dos objetos en los dos agrupamientos.

$$d(r,s) = \max(dist(x_{ri}, x_{sj}), i \in (i,\dots,n_r), j \in (1,\dots,n_s) \qquad \text{(AIII.2)}$$

c. Vinculación promedio, utiliza la distancia promedio entre todos los pares de objetos en cualquiera de los dos agrupamientos.

$$d(r,s) = \frac{1}{n_r n_s}\sum_{i=1}^{n_r}\sum_{j=1}^{n_s} dist(x_{ri}, x_{sj}) \qquad \text{(AIII.3)}$$

d. Vinculación centroide, utiliza la distancia Euclídea entre los centroides de los dos agrupamientos.

$$d(r,s) = \left\|\overline{x}_r - \overline{x}_s\right\|_2 \qquad \text{(AIII.4)}$$

donde $\overline{x}_r = \frac{1}{n_r}\sum_{i=1}^{n} x_{ri}$ y $\| \ \|_2$ se refiere a la distancia Euclídea.

e. Vinculación media, utiliza la distancia Euclídea entre los centroides ponderados de los dos agrupamientos,

$$d(r,s) = \left\|\tilde{x}_r - \tilde{x}_s\right\|_2 \qquad \text{(AIII.5)}$$

donde $\tilde{x}_r$ y $\tilde{x}_s$ son los centroides pesados para los agrupamientos r y s. Si el agrupamiento r fue creado por la combinación de los agrupamientos p y q, $\tilde{x}_r$ es definido recursivamente como

$$\tilde{x}_r = \frac{1}{2}(\tilde{x}_p + \tilde{x}_q)$$

f. Vinculación de Ward, utiliza la suma incremental de los cuadrados, es decir, el incremento en la suma total de los cuadrados dentro del agrupamiento, como resultado de la unión de dos grupos. La suma de los cuadrados dentro del agrupamiento es definida como la suma del cuadrado de las distancias entre todos los objetos en el agrupamiento y el centroide del agrupamiento. La distancia equivalente es:

$$d^2(r,s) = n_r n_s \frac{\|\bar{x}_r - \bar{x}_s\|_2^2}{(n_r + n_s)} \qquad \text{(AIII.6)}$$

g. Promedio ponderado de vinculación, utiliza una definición recursiva para la distancia entre dos agrupamientos. Si el agrupamiento r fue creado mediante la combinación de los agrupamientos p y q, la distancia entre r y otro agrupamiento s se define como el promedio de las distancias entre p y s y la distancia entre q y s:

$$d(r,s) = \frac{\left(d(p,s) + d(q,s)\right)}{2} \qquad \text{(AIII.7)}$$

IV. Eliminación de Mínimos Locales y Descripción del Cálculo Iterativo en el Método K-Medias.
Eliminación de mínimos locales

Al igual que sucede en muchos otros problemas de optimización numérica, la solución que se alcanza con el método K-Medias depende a menudo del punto de partida, en este caso la posición inicial del centroide de cada agrupamiento. Es posible así alcanzar un mínimo local, donde la reasignación de cualquier punto a un nuevo agrupamiento debería incrementar la suma total de distancias centroide-punto, pero donde puede existir realmente una mejor solución. Para solucionar este problema, es posible especificar en el método el número de 'réplicas', es decir, el número de veces en que se repetirá el proceso de agrupación, cada uno con un nuevo conjunto de posiciones iniciales del centroide del agrupamiento. Por supuesto, la mejor solución será aquella para la cual la suma de las distancias centroide-punto para cada uno de los agrupamientos sea mínima.

Descripción del algoritmo

El algoritmo consta de dos partes:

Primera fase. Cada iteración consiste en la reasignación colectiva de elementos al centroide del agrupamiento más cercano, todos a la vez, seguida de un nuevo cálculo de las posiciones de los centroides. La primera fase ocasionalmente converge a soluciones que son un mínimo local; es más probable alcanzar un mínimo global si se trabaja con pequeños grupos de datos. La fase de actualización colectiva es rápida, pero posiblemente sólo aproxime una solución que sea el punto de partida de la segunda fase.

Segunda fase. Los elementos son reasignados individualmente si con ello se reduce la suma de distancias, y en cada reasignación se calcula la ubicación del centroide del agrupamiento. Cada iteración consiste en la ubicación de todos los elementos. En esta fase la solución converge a un mínimo local, aunque puede haber otro mínimo local con menor suma total de distancias. Generalmente el problema de hallar un mínimo global puede ser resulto únicamente por medio de una selección exhaustiva de los puntos de partida, aunque la utilización de varias réplicas con puntos de partida aleatorios generalmente converge a una solución que es un mínimo global.

V. Definición del Parámetro Silueta

La definición de los valores silueta es la siguiente: dado un grupo G_1 y un objeto i asignado a éste, la disimilitud promedio de i para todos los objetos j en G_1 está dada por:

$$a(i) = \sum_j d(i, G_{1j}) / \text{número de objetos en } G_1 \qquad \text{(AV.1)}$$

donde $d(i, G_{1j})$ es la distancia de cada objeto i a cada objeto j en G_1. La menor disimilitud correspondiente a i respecto a cualquier otro agrupamiento, $b(i)$, también es calculada. Si i es más similar a los objetos en un grupo G_2 que a los del grupo G_1, entonces:

$$b(i) = \sum_j d(i, G_{2j}) / \text{número de objetos en } G_2 \qquad \text{(AV.2)}$$

Por lo tanto, el valor silueta $s(i)$ definido para un objeto i es:

$$s(i) = \left[b(i) - a(i)\right] / \max\left\{a(i), b(i)\right\} \qquad \text{(AV.3)}$$

donde $\max\left\{a(i), b(i)\right\}$ se refiere al mayor valor entre $a(i)$ y $b(i)$.

VI. El Clasificador del Método K-Vecinos Más Cercanos

Si se quiere conocer la clase a la que pertenece un objeto dado, entre varias clases posibles, se introduce el concepto de clasificador. Un clasificador es una función que asigna un objeto a una clase determinada, para lo cual se basa en el conocimiento de sus variables atributo. Existen dos tipos de clasificadores:

- Paramétricos: asumen que la distribución estadística que sigue el conjunto de variables es conocido, y trata de estimar los parámetros de dicha distribución.
- No-paramétricos: no asume ninguna distribución en particular. El clasificador se construye únicamente con los datos del conjunto de entrenamiento.

Entre los clasificadores no-paramétricos, el más conocido es el basado en el método K-vecinos más cercanos: si K_i es el número de objetos que pertenecen a la clase G_i entre los K vecinos más cercanos al objeto considerado x, la probabilidad a posteriori P (G_i | **d**) (la probabilidad de que la clase sea G_i cuando x se describe con el conjunto de variables **d**) se estima como $\frac{K_i}{K}$.

De esta manera, el clasificador asigna x a la clase más frecuente entre sus K vecinos más cercanos, según una cierta medida de distancia.

VII. Algunos Algoritmos Utilizados en Matlab

Algoritmo clusterskmeans.m

```
function [Resultkmeans] = clusterskmeans(p,tot,nclusters,percent)
% INPUT:
% p = experimental property
% tot = data matrix with pool of descriptors
% nclusters = number of clusters to be created
% percent = percent of total compounds to be used as part of train and
%                     val sets
% OUTPUT:
% Resultkmeans = returns in each row in the following order:
%        the descriptor number, mean silhouette value, minimum silhouette
%        value, rmse(train), rmse(val), rmse(test), rmse(train)-mse(val),
%        the within-cluster sums of point-to-centroid distances, the
%        sum of within-cluster sums of point-to-centroid distances, and
%        the number of objects in each cluster
% Pablo R. Duchowicz
% INIFTA, La Plata, Argentina
% Created: 2nd March 2011
warning off
resj=[];
[d1, d2]=size(tot);
```

```
for j=1:d2
    apt=0;
    vecto=[j];
    [IDX0, C, sumd, D, silh3, msilh3, ncomp] = kmca(tot(:,vecto), nclusters, 'distance', 'sqEuclid-
ean', 'off', 50);
    if msilh3==-999
        apt=1;
    else
        if msilh3==1
            apt=1;
        else
            if silh3>0
                apt=0;
            else
                apt=1;
            end
        end
    end
    if apt==0
        res=sumd';
        [Idmol Ncomp] = idmol(IDX0, nclusters);
        vari=0;
        for k=1:nclusters
            if Ncomp(k)==0
                vari=1;
            end
        end
        if vari==0
            [train, val, test] = analysis2(Idmol,percent);
            [Res] = rmsr(p(train), [1], tot0(train,:), 1);
           [rsstrain rssval rsstest] = trainvaltestrss(train,val,test,p,tot0,Res(2));
          resj=[resj;[vecto,msilh3,min(silh3),Res(2),rsstrain,rssval,rsstest,abs(rsstrain-rssval),
(1-rsstrain/rssval)*100, res,sum(res),ncomp]];
        end
    end
end
if size(resj,1)>0
    resj=sortrows(resj,(dv+9));
end
Resultkmeans=resj;
```

Algoritmo clustersknn.m

```
function  [Resultknn] = clustersknn(p,tot,nclusters,percent)
% INPUT:
% p = experimental property
% tot = data matrix with pool of descriptors
% nclusters = number of clusters to be created
% percent = percentage of total compounds to be used as part of train and
%           val sets
% OUTPUT:
% Resultknn = returns in each row the results in the following order:
%        the classifying descriptor's number, the best descriptor of
%        linear model, rmse(train), rmse(val), rmse(test), rmse(train)-
%        rmse(val), the percentage difference between train and val, and
%        the number of objects in each cluster
% Pablo R. Duchowicz
% INIFTA, La Plata, Argentina
% Created: 15th March of 2011
warning off
resj = [];
[d1, d2] = size(tot);
for j=1:d2
    vecto = [j];
    [CN] = centers(tot(:,vecto),nclusters);
    c = knnclassify(tot(:,vecto),CN,(1:nclusters)');
    [Idmol Ncomp] = idmol(c, nclusters);
    vari=0;
    for k = 1:nclusters
        if Ncomp(k)==0
            vari=1;
        end
    end
    if vari == 0
        [train, val, test] = analysis2(Idmol,percent);
        [Res] = rmsr(p(train), [1], tot(train,:), 1);
        [rsstrain rssval rsstest] = trainvaltestrss(train,val,test,p,tot,Res(2));
                resj=[resj;[vecto,Res(2),rsstrain,rssval,rsstest,abs(rsstrain-rssval),(1-rsstrain/
rssval)*100,Ncomp]];
    end
end
if size(resj,1)>0
    resj=sortrows(resj,(dv+7));
end

Resultknn=resj;
```

EX DIRECTORES DE LOS ANALES DE LA SOCIEDAD CIENTÍFICA ARGENTINA (*)

Ing. Pedro Pico
Ing. Luis A. Huergo
Dr. Carlos Berg
Dr. Estanislao S. Zeballos
Ing. Eduardo Aguirre
Ing. Carlos Bunge
Dr. Angel Gallardo
Dr. Félix F. Outes
Dr. Horacio Damianovich
Ing. Julio R. Castiñeiras
Ing. Emilio Rebuelto
Ing. José S. Gandolfo
C. de Nav. Emilio L. Díaz
Dr. Pedro Cattáneo
Ing. Guillermo White
Dr. Valentín Balbín
Ing. Luis A. Viglione
Dr. Carlos María Morales
Ing. Jorge Duclout
Ing. Miguel Iturbe
Ing. Domingo Nocetti
Ing. Santiago Barabino
Dr. Eduardo Carette
Dr. Claro D. Dassen
Ing. Alberto Urcelay
Dr. Reinaldo Vanossi
Dr. Andrés O. M. Stoppani
Dr. Eduardo A. Castro
Dr. Alfredo Kohn Loncarica

(*) Desde 1876 a 1902: Presidente de la Comisión Redactora.

PRESIDENTES HONORARIOS DE LA SOCIEDAD CIENTÍFICA ARGENTINA

1.- Prof. Dr. Andrés O. STOPPANI. † (1915 - 2003)

2.- Dr. Carlos Pedro BLAQUIER. (1927)

LA REVISTA
ANALES DE LA SOCIEDAD CIENTÍFICA ARGENTINA
HA SIDO INCLUIDA EN LA BASE DE DATOS

LATINDEX

(Directorio y Catálogo)
www.latindex.unam.mx

Instituto de Cursos y Conferencias
Dr. Carlos Pedro Blaquier
Sociedad Científica Argentina

Programación 2011

Cursos de verano

A propósito de la metáfora del libro del mundo
A cargo de: Dr. José González Ríos

Los fundamentos filosóficos del liberalismo moderno
A cargo de: Dr. Martín D'Ascenzo

Ciclo Anual

El Universo Obscuro
Dr. Héctor Vucetich

Participación política y reforma electoral
Dr. Alberto Dalla Vía

Para una historia del problema del lenguaje
Dr. José González Ríos

La Constitución y los Sistemas Electorales
Dr. Jorge Reinaldo Vanossi

Lógica Clásica y Otras Lógicas
Dra. Gladys Palau

Los Materiales del Arte
Dr. José Selles Martínez

Procreación Asistida – Aspectos Jurídicos
Dr. Augusto C. Belluscio

Borges, el hombre y las letras
Prof. Alicia Ardila

Estrategia – Pensar la acción
Dr. José Luis Speroni

La Evolución de las Ideas
Ing. Roberto Cook

Enfrentar La Tempestad - Para una lectura de la última obra de Shakespeare
Mg. Mónica Maffía

Eduardo III – Para una lectura de la nueva obra de Shakespeare
Mg. Mónica Maffía

Medios digitales
en la era de las comunicaciones
Dra. Alicia Jubert

ABIERTA LA INSCRIPCIÓN
2011

Se entregan certificados de asistencia
en todos los cursos.

INSCRIPCIONES

Av. Santa Fe 1145 Ciudad de Buenos Aires

Más información: Te. 4816-4745/4816-5406
Email: cursos@cientifica.org.ar

INSTRUCCIONES PARA LOS AUTORES

Las siguientes *Instrucciones para los autores* constituyen el reglamento de publicaciones de los ANALES DE LA SOCIEDAD CIENTIFÍCA ARGENTINA.

1) Generales

Los ANALES DE LA SOCIEDAD CIENTÍFCA ARGENTINA constituyen una revista multidisciplinaria, fundada en 1876, que considera para su publicación trabajos de cualquier área de la ciencia.

Los originales deben ser enviados al director, a Av. Santa Fe 1145, Buenos Aires, CP.:1059, República Argentina, en tres copias en papel, a dos espacios, tamaño carta, acompañados de su correspondiente disquete. Los disquetes deberán estar rotulados con el nombre del autor o del primer autor si son varios haciendo constar el sistema computacional usado para grabar el mismo, el tipo y versión del procesador utilizado y nombres de los archivos.

Los autores serán notificados de inmediato de la recepción de sus originales. Dicha notificación no implica la aceptación del trabajo. Los originales son enviados a uno o más 'arbitros, quienes asesoran al director y a la comisión de redacción acerca de la aceptación, rechazo o sugerencia de modificaciones. La decisión final respecto a la publicación o no del trabajo es solamente responsabilidad del director.

Los originales remitidos para su publicación en los ANALES deben ser inéditos y no hallarse en análisis para su publicación en otra revista o cualquier otro medio editorial.

Todo trabajo aceptado en los ANALES no podrá ser publicado en otro medio gráfico sin previo consentimiento de la dirección.

Los ANALES se reservan el derecho de rechazar sin más trámite a aquellos originales que no se ajusten a las normas expuestas en la presente guia de *Instrucciones para los autores*.

Los ANALES constan de las siguientes secciones:

-artículos de investigación
-notas breves de investigación
-articulos de revisión y/o actualización
-editoriales
-recensiones
-cartas a la dirección
-informaciones del quehacer de la SOCIEDAD CIENTIFICA ARGENTINA
-informaciones científicas y acádemicas de interés general

Los autores, al remitir sus trabajos, deberán hacer constar la sección, a la que según su juicio, corresponden sus aportes y consignar claramente la dirección postal, teléfono, fax y dirección electrónica (si la tuviere) a la cual se remitirá toda información coreeniente al original.

2) Originales

Los ANALES DE LA SOCIEDAD CIENTIFICA ARGENTINA publicarán trabajos escritos en los idiomas: español, francés, inglés y portugués.

Los originales deberán respetar la siguiente estructura:

1ª pagina:

-Título del trabajo: no mayor de veinticinco (25) palabras
-Nómina de los autores, institución o instituciones a la que pertenecen cada uno de ellos.
-Institución en la que se llevó a cabo el trabajo en el caso que difiera de la institución de pertenencia.
-Domicilio postal y electrónico (si lo tuviere)

2ª página:

-Resumen en idioma español de no más de 400 palabras, con su correspondiente traducción al inglés. La traducción al inglés deberá incluir el título del trabajo cuando éste haya sido escrito en español y viceversa, si el trabajo se halla escrito en inglés el resumen en español deberá incluir la traducción del título.
-La inclusión de resumenes en francés y portugués es facultativa de los autores.
-Palabras claves para el registro bibliogáfico e inserción en bases de datos, en español e inglés.

En las páginas siguientes se incluirán las secciones Introducción, Materiales y Métodos, Resultados, Discusión, Agradecimientos y Referencias. A continuación se agregarán las tablas con sus títulos, leyendas de las figuras y gráficos y finalmente las figuras y gráficos preparados como se indica más abajo.

El tipeado del manuscrito deberá hacerse a doble espacio en papel tamaño carta (aprox. 21 cm x 29cm), dejando 3 cm de márgenes izquierdo, superior e inferior, debiéndose numerar secuencialmente todas las páginas.

No se aceptará la inserción de notas de pie de página. Cuando ello sea necesario, se deberá incluir tales notas en el mismo texto.

Se recomienda emplear el Sistema Métrico Decimal de medidas y las abreviaturas universales estándar.

Solo se permitirá el empleo del Sistema Internacional de Unidades para las medidas.

Como regla general no se deberá repetir la misma información en tablas, figuras y texto. Salvo en casos especiales que justifiquen alguna excepción se aceptará presentar esencialmente la misma la información en dos formas simultáneas.

Cada sección se numerará consecutivamente, recomendándose no emplear subsecciones.

3) Tablas

Las tablas deben prepararse en hojas aparte y a doble espacio. Las mismas incluirán un título suficientemente aclaratorio de su contenido y se indicarán en el texto su ubicación, señalándolo con un lápiz sobre el margen izquierdo.

Cada tabla se numerará consecutivamente con números arábigos. Solo se deberá incluir en las tablas información significativa, debiéndose evitar todo dato accesorio y/o que pueda ser mejor informado en el mismo texto del trabajo.

Cada tabla se tipeará en hoja separada.

Los títulos de las filas y las columnas deben ser lo suficientemente explícitos y consistentes, pero al mismo tiempo se recomienda concisión en su preparación.

4) Ilustraciones

Las ilustraciones (gráficos y fotografías) deberán ser de suficiente calidad tal que permitan una adecuada reproducción debiéndose tener en cuenta que la reproducción directa de los mismos conlleva una relación entre 1:2 y 1:3. Todas las ilustraciones se numerarán consecutivamente y en el reverso de las mismas se indicarán con lápiz blando el nombre de los autores, el número de la misma y cuando corresponda la orientación para su pertinente impresión.

Los títulos de las ilustraciones se tipearán en hoja aparte, debiéndose denotar el posicionado de las mismas en el texto por medio de una indicación con lápiz en el margen izquierdo.

Las dimensiones de las ilustraciones no deberán exceder las de las hojas del manuscrito y no se deberán doblar.

Los gráficos se dibujarán con tinta china sobre papel vegetal de buena calidad y por los mismos medios se incluirán los símbolos, letras y números correspondientes. No se deberá tipear símbolo, letra o número alguno en los gráficos y fotografías.

Enviar un original y dos copias de cada ilustración. Las fotografías solo se podrán enviar en blanco y negro, ya que que no es posible imprimir fotografías en otros colores.

Cada ilustración se presentará en hoja separada.

5) Referencias

Los ANALES adoptan el sistema de referencias por orden, el cual consiste en citar los trabajos en el orden que aparecen por medio de número cardinal correspondiente. Los libros se indicarán en la lista de referencias citando el/los autor/es, título, edición, editorial, ciudad, año y página inicial. Para indicar capítulo de libro se añadirá a lo anterior el título del mismo y el nombre del editor.

El listado de referencias se tipeará en hoja separada y a doble espacio. Se recomienda especialmente a los autores emplear las abreviaturas estándar sugeridas por las propias fuentes.

Solo se admitirán citas de publicaciones válidas y asequibles a los lectores por los medios normales debiéndose evitar recurrir a informes personales, tesis, monografías, trabjos en prensa, etc., de circulación restringida.

Lo que sigue son algunos ejemplos de citas bibliográficas en la lista de referencia:

Publicación periódica: A. M. Sierra y F. S. Gonzalez, J. Chem. Phys. 63 (1977) 512.

Libro: R. A.. Day, How to write and publish a Scientific paper, Second Edition, ISI Press, Philadelphia, 1983, p 35.

Capítulo del libro: Z. Kaszbab, Family Tenebrionodae en W. Wittmer and Buttiper (Eds.) Famma of Saudi Arabia, Ciba-Geigy, Basel, 1981, p3-15.

Conferencia o Simposio: A. Ernest, Energy conservation measures in Kuwait buildings. Proceedings of the First Symposium on Thermal Insulation in the Gulf States, Kuwait Institute for Scientific Research, Kuwait, 1975, p 151.

Se recomienda revisar cuidadosamente las citas en el texto y la lista de referencias a los efectos de evitar inconsistencias y/u omisiones.

Pruebas: todo artículo deberá ser revisado en la forma de prueba de galera por el autor indicado en la carta de presentación del trabajo, la cual se devolverá debidamente corregida a las 72 horas de recibida a la redacción de los ANALES. No se admitrá en forma alguna alteración sustancial del texto y en caso imprescindible se procederá a la inclusión al final del trabajo de lo que correspondiera bajo el título de " Nota agregada en la prueba".

ANALES DE LA SOCIEDAD CIENTÍFICA ARGENTINA

Organo de la Sociedad Científica Argentina.

Revista fundada el 14 de diciembre de 1875, cuyo primer número apareció el 14 de enero de 1876.

Se viene editando continuadamente desde esta fecha.

Director

Dr. Angel Alonso

Comisión de Redacción

Dra. María H. Bertoni

Dr. Alberto Boveris

Dr. Horacio H. Camacho

Dr. Eduardo Castro

Dra. Stella M. González Cappa

Dr. Gabriel A. Gutkind

Dra. Georgina R. de Lores Arnaiz

Dr. Federico Pérgola

Dr. Eduardo Antonio Pigretti

Dr. Humberto Quiroga Lavié

Dr. Rodolfo P. Rothlin

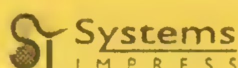
Ing. Juan J. Sallaber

Dr. Daniel Sordelli

Dr. Jorge Reinaldo Vanossi

Dr. Pedro Yañez

Editado por:



Uruguay 827 - Capital Federal - stms@fibertel.com.ar

Buenos Aires, Enero 2012

ISSN 0037-8437

ANALES

DE LA

SOCIEDAD CIENTIFICA ARGENTINA

AÑO 2012 - VOLUMEN 246 - N° 1

SUMARIO Pág.